Title
Essays on Search Frictions in Financial Markets

Author
Uslu, Semih

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Search Frictions in Financial Markets

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Economics

by

Semih Uslu

2016

ABSTRACT OF THE DISSERTATION

Essays on Search Frictions in Financial Markets

by

Semih Uslu

Doctor of Philosophy in Economics

University of California, Los Angeles, 2016

Professor Pierre-Olivier Weill, Chair

This dissertation consists of three chapters about search frictions in financial markets.

Chapter 1: "Pricing and Liquidity in Decentralized Asset Markets"

I develop a search-and-bargaining model of liquidity provision in over-the-counter markets where investors differ in their search intensities. A distinguishing characteristic of my model is its tractability: it allows for heterogeneity, unrestricted asset positions, and fully decentralized trade. I find that investors with higher search intensities (i.e., fast investors) are less averse to holding inventories and more attracted to cash earnings, which makes the model corroborate a

number of stylized facts that do not emerge from existing models: (i) fast investors provide intermediation by charging a speed premium, and (ii) fast investors hold larger and more volatile inventories. I also calibrate the model, demonstrate that it produces realistic quantitative outcomes, and use it to study the effect of trading frictions on the supply and price of liquidity. The results have policy implications concerning the Volcker rule.

Chapter 2: "Price Dispersion and Trading Activity during Turbulent Times"

I construct a dynamic model of crises in a decentralized asset market that operates via search and bargaining. The crisis is modeled as a one-time aggregate shock to uncertainty with a random recovery. The arrival of the crisis shock leads to an increase in both the volatility of asset payoff and the volatility of investors' background risk. The equilibrium path for investors' valuations, terms of trade, and the distribution of investors' positions is characterized in closed form both during the crisis and during the recovery. Tractability of the model allows me to derive natural proxies for price dispersion and trading activity. I show that both volatility of asset payoff and volatility of background risk contribute to higher level of price dispersion during the crisis. Trading activity might be higher or lower depending on the increase in the volatility of background risk relative to the increase in the volatility of asset payoff, consistent with the "flight-to-quality" observations during extreme episodes. A flight to the asset market always starts with a "heating-up" in trading activity but a flight from the market might start with a dry-up or heating-up during the onset of the crisis. If the relative increase in the volatility of asset payoff is too high, a period of fire sales is triggered leading to a short heating-up before the complete dry-up of the trading activity. I calibrate the model according to the U.S. corporate bond market data and show that it captures the observations during the subprime crisis.

Chapter 3: "Endogenous Liquidity and Cross-section of Returns in Dynamic Bargaining Markets"

The empirical analysis of liquid/illiquid asset pairs reveals the existence of a return differential (liquidity premium) between those types of assets. The time variation in liquidity premia is delineated by the term "flight-to-liquidity," meaning that liquidity premia are higher during extreme market episodes. In this paper, I extend the search-and-bargaining model of Weill (2008) by allowing for risk aversion, to explain this observation. Risk-averse investors optimally allocate their limited budgets of search efforts to various assets. This extension allows me to examine the relationship between risk and liquidity of assets in the cross-section and over time. My model generates endogenous cross-sectional liquidity differentials corroborating much of the empirical evidence. Furthermore, I show that when asset payoffs are more volatile, trade surpluses are higher because idiosyncratic hedging quality differentials are wider. Higher trade surpluses lead to higher value of search, and in turn, higher opportunity cost of committing to a particular asset, especially to an illiquid one. Therefore, periods of high volatility are associated with a flight-to-liquidity.

The dissertation of Semih Uslu is approved.

Daniel Dumitru Andrei

Andrew Granger Atkeson

Simon Adrian Board

Pierre-Olivier Weill, Committee Chair

University of California, Los Angeles

2016

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Chapter 1: I am deeply indebted to my advisor, Pierre-Olivier Weill, for his continued guidance, his encouragement, many detailed comments, and suggestions. I am grateful to Darrell Duffie for his comments, which improved the paper. I also would like to thank, for fruitful discussions and comments, Daniel Andrei, Andrew Atkeson, Simon Board, Bingen Cortazar, Adrien d'Avernas, Ilker Kalyoncu, Fernando Martin, Bugra Ozel, Marek Pycia, Paulina Restrepo-Echavarria, Guillaume Rocheteau, Tomasz Sadzik, Guner Velioglu, Christopher Waller, Stephen Williamson, and participants in the various seminars at UCLA Economics, Anderson Finance, St. Louis Fed, the UCI Search and Matching PhD Workshop, and the Chicago Fed/St. Louis Fed Summer Money Workshop. Part of this project was completed when I was a Dissertation Intern at the Federal Reserve Bank of St. Louis. I gratefully acknowledge their hospitality. The views expressed in this paper are my own and do not necessarily reflect those of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

Chapter 2: I am grateful to Pierre-Olivier Weill for his guidance and support.

Chapter 3: I am grateful to Pierre-Olivier Weill for his guidance and support. I also want to thank Daniel Andrei, Andrew Atkeson, and Simon Board for their comments and suggestions.

VITA

2010-2011	Junior Teaching Assistant, Economics, Sabanci University, Istanbul, Turkey
2011	Bachelor of Arts in Economics, Sabanci University. *with highest honors*
2012-2016	Teaching Assistant, Department of Economics, University of California, Los Angeles
2012	Master of Arts in Economics, University of California, Los Angeles
2013-2014	Research Assistant for Prof. Pierre-Olivier Weill, University of California, Los Angeles
2013	Candidate of Philosophy in Economics, University of California, Los Angeles
2014	Research Assistant for Prof. Pierre-Olivier Weill, National Bureau of Economic Research, Cambridge, Massachusetts
2015	Dissertation Intern, Federal Reserve Bank of St. Louis, St. Louis, Missouri

CHAPTER 1

Pricing and Liquidity in Decentralized Asset Markets

1 Introduction

Recent empirical analyses of over-the-counter (OTC) markets point to a high level of heterogeneity among intermediaries with respect to transaction frequency, terms of trade, and inventories.[1] Some intermediaries appear to be central in the network of trades: They trade very often, and hold large and volatile inventories. Moreover, they face systematically different terms of trade. In the municipal bond market, for example, central intermediaries earn higher markups compared to peripheral intermediaries.[2] On the other hand, central intermediaries in the market for asset-backed securities earn lower markups.[3] In this paper, I provide a theoretical model that captures the economic incentives of intermediaries which give rise to these empirical trading patterns.

More precisely, I consider an infinite horizon dynamic model in which investors meet in pairs to trade an asset. I go beyond the literature by considering investors who can differ in their search intensities, time-varying hedging needs, and asset holdings. I provide an analytical characterization of the steady state equilibrium that includes the distribution of asset holdings, bilateral trade quantities, and prices. The rich heterogeneity in the model allows me to reproduce the observed trading patterns in OTC markets, and, therefore, provides a natural laboratory for policy analysis. In a calibrated example of my model, I show that,

[1] The heterogeneity among intermediaries is documented for the municipal bond market (Li & Schürhoff, 2012), the fed funds market (Bech & Atalay, 2010), the overnight interbank lending market (Afonso, Kovner & Schoar, 2014), the market for asset-backed securities (Hollifield, Neklyudov & Spatt, 2014), and the market for credit default swaps (Siriwardane, 2015).

[2] See Li and Schürhoff (2012).

[3] See Hollifield et al. (2014).

in markets where central intermediaries earn higher markups, the further concentration of intermediation activity in the hands of these central intermediaries is beneficial for social welfare, while it is harmful in markets where central intermediaries earn lower markups. This suggests that the empirical relationship between markups and centrality helps predict the potential effects of regulatory actions, such as the Volcker rule and MiFID I/II, which aim at reducing the concentration of intermediation activity.

In my model, intermediation arises endogenously as a result of the interaction of investor heterogeneity and search frictions. I model heterogeneity in search intensity among investors as heterogeneity in the number of trading specialists with whom the investors are endowed. Specialists randomly contact each other to trade a risky asset on behalf of investors. Thus, in effect, investors with higher number of specialists have higher search intensities. Conditional on a contact, both price and quantity are determined endogenously by bilateral bargaining. Importantly, the quantity traded is endogenous since I do not impose the usual $\{0, 1\}$ holding restriction of the literature. This generalization allows me to analyze how financial intermediaries optimally manage their inventories' sizes and facilitate trading.

The model can rationalize the trading patterns observed in OTC markets: namely, the heterogeneity across intermediaries in transaction frequency, terms of trade, and inventories. I show that "fast investors" (who have higher search intensities) have relatively stable marginal valuations that are close to the average marginal valuation of the market, so they become endogenously central. Therefore, as observed in the data, fast investors hold larger and more volatile inventories to provide intermediation to slow investors. In return, these fast investors charge a speed premium as the price of the liquidity they provide. I show that the relationship between the centrality of an investor and the intermediation markups she earns arises as a result of two competing effects: stable marginal valuations and speed premium. Her stable marginal valuations tend to reduce the markups she charges, by making inventory-holding less risky. If this is the dominant effect, we observe a negative relationship between centrality and markups. When the speed premium is dominant, we observe a positive relationship between centrality and markups. I find that the speed premium is dominant when search frictions are severe or investors experience liquidity shocks very frequently.

The main analytical difficulty posed by this model is keeping track of the endogenous joint distribution of asset holdings, hedging needs, and search intensities. However, using convolution methods, I show that marginal valuations, terms of trade, and the first conditional moment of equilibrium distribution can be found in closed form up to effective discount rates that solve a functional equation, so that the analysis remains relatively tractable. I also provide a recursive characterization of higher order conditional moments of the equilibrium distribution. Therefore, one contribution of this paper to the literature is methodological: It drops the restrictions on asset positions, without forgoing the investor heterogeneity or fully decentralized trading structure. With this level of generality, my model offers a unified framework to address positive and normative issues surrounding OTC markets.

The main mechanism behind different trading behaviors of fast and slow investors is that heterogeneity in search intensities leads to heterogeneous *effective discount rates* at which investors discount their current utility flow. The effective discount rate is higher for fast investors because they are able to rebalance their holdings faster. This increases the importance of the option value of search, and decreases the importance of the current utility flow from holding the asset. In other words, high effective discount rates lead to the lower sensitivity of marginal valuations to asset holdings. Therefore, fast investors put less weight on their asset positions and more weight on their cash earnings when bargaining with counterparties. Each bilateral negotiation results in a trade size that is more in line with the slower counterparty's hedging need and a trade price that contains a premium benefitting the faster counterparty. Controlling for the level of marginal valuation, fast investors provide more intermediation due to this effective discount rate channel. In addition, fast investors engage in higher simultaneous buying and selling activity due to the higher intensity of matching with counterparties. However, the effective discount rate channel leads to an increase in the intermediation level above and beyond that direct effect. As in the data, not only do fast investors trade more often, but they also trade larger quantities on average, in each match.

Another important result of my model is that investor heterogeneity makes the equilibrium constrained inefficient due to a hold-up problem typical of *ex post* bargaining environments. The root cause of inefficiency is the price impact. When negotiating for a trade quantity, investors recognize the fact that their trades will create a price impact in the future and that the price impact is increasing in the surplus that those future trades generate. Consequently, at the margin, investors tend to take more cautious positions than is socially optimal which would lead to larger price impacts in future trades. In a calibrated example, I show that the welfare loss caused by OTC market frictions can be as large as 4% of the constrained efficient welfare in consumption equivalent terms.[4] This result reveals that there is room for beneficial intervention in markets with *ex post* bargaining and investor heterogeneity, which are virtually all OTC markets. For the inefficiency result, investor heterogeneity in hedging need or speed is essential. Afonso and Lagos (2015) show that if there is no investor heterogeneity, the equilibrium of a fully decentralized market with unrestricted holdings is constrained efficient, even though there is a hold-up problem. Because all investors are identical in their exogenous characteristics, their marginal valuations are distorted in exactly the same way, so the negotiated trade quantities coincide with the planner's quantities.

Finally, I present a calibrated numerical example that demonstrates that my model can produce quantitatively meaningful results in terms of distribution of trade sizes and the relationship between degree centrality and intermediation markups. I, then, use this calibrated model to conduct comparative statics analysis. Specifically, I analyze how a change in the

[4] The welfare notion I use is *ex ante* welfare, which is defined as the sum of all investors' certainty equivalents at date 0.

central intermediaries' search intensities affects the welfare. Investors trade off between the benefit of hedging and the cost of risk-bearing when they invest in the asset. An increase in the main intermediaries' search intensities causes the further concentration of intermediation activity in the hands of those main intermediaries and, in turn, leads to a higher hedging benefit and a higher cost of risk-bearing at the same time. If search frictions are severe or investors experience liquidity shocks very often, the increase in hedging benefit becomes dominant, and we observe an increase in welfare. Otherwise, the cost of risk-bearing becomes dominant, and we observe a decline in welfare. This result relates the welfare impact of concentration to the sign of the relationship between centrality and markups. In markets with a positive relationship between centrality and markups (e.g. municipal bond market) the impact of an increase in fast investors' search intensities on social welfare turns out to be positive, while it is negative in markets with a negative relationship between centrality and markups (e.g. the market for asset-backed securities).

These results inform the debate on the effects of a section of the Dodd-Frank Act, often referred as "the Volcker rule," which bans proprietary trading by banks and their affiliates. It is commonly agreed that the Volcker rule effectively reduces the ability of intermediaries to provide liquidity.[5] Accordingly, in my model, I capture this in a stylized way by decreasing the number of specialists the central intermediaries have. My model predicts different welfare impacts for different markets. While it would be beneficial for markets with negative relation between centrality and markups, it would be harmful for markets with positive relation between centrality and markups.

1.1 Related Literature

A fast-growing body of literature, spurred by Duffie, Gârleanu, and Pedersen (2005), has recently applied search-theoretic methods to asset pricing. The early models in this literature,

[5] See Duffie (2012b).

such as Duffie, Gârleanu and Pedersen (2007), Weill (2008), and Vayanos and Weill (2008),[6] studied theories of fully decentralized markets in a random search and bilateral bargaining environment and used these theories to present a better understanding of the individual and aggregate implications of distinctively non-Walrasian features of those markets. These models maintain tractability by limiting the investors to two asset positions, 0 or 1. Another part of this body of literature, with papers by Gârleanu (2009) and Lagos and Rocheteau (2007, 2009), eliminates the $\{0, 1\}$ restriction on holdings by introducing a partially centralized market structure.[7] In their framework, investors are able to trade in a centralized market but only infrequently and by paying an intermediation fee to exogenously designated dealers who have continuous access to the centralized market. These models show that investors' decisions at the intensive margin provide them with the flexibility to respond to changes in market conditions.

My model is the first model that combines unrestricted asset holdings, fully decentralized market structure and heterogeneity in search intensities. The combination of unrestricted holdings and fully decentralized trade is essential for the analysis I conduct because fully decentralized trade is necessary for endogenous intermediation, and unrestricted holdings are necessary for the study of optimal inventory holding behavior. To the best of my knowledge, there are two papers with this combination. Afonso and Lagos (2015) study trading dynamics in the Fed Funds market. In their model, banks are homogeneous in terms of preferences and search intensities. The basic insight from their model on "endogenous intermediation" applies to my model as well. They show that banks with average asset holdings endogenously become "middlemen" of the market by buying from banks with excess reserves and selling

[6] The framework of Duffie et al. (2005) has also been adopted to analyze a number of issues, such as market fragmentation (Miao, 2006), clientele effects (Vayanos & Wang, 2007), the congestion effect (Afonso, 2011), commercial aircraft leasing (Gavazza, 2011a), and the co-existence of illiquid and liquid markets (Praz, 2014).

[7] Other papers that use the same trading framework include Lagos, Rocheteau, and Weill (2011), Lester, Rocheteau, and Weill (2015), Pagnotta and Philippon (2015), and Randall (2015).

to banks with low reserves. Relative to Afonso and Lagos (2015), my contribution is to solve for a stochastic steady-state with two new dimensions of heterogeneity: hedging need and search intensity. As I explain above, these are important for explaining stylized OTC market facts and obtaining new normative results. Cujean and Praz (2015) study the impact of information asymmetry between counterparties. Although their model also features unrestricted asset holdings and a fully decentralized market structure, my work is different from theirs in that they assume all investors have the same search intensity. In order to analyze the microstructure of OTC markets, I introduce search heterogeneity but keep the usual symmetric information assumption of the literature. Then, I study the resulting topology of trading relations.

My paper is also related to the literature on the trading networks of financial markets. Recent works include Babus and Kondor (2012), Farboodi (2014), Gofman (2011), and Malamud and Rostek (2012). Atkeson, Eisfeldt, and Weill (2015), Chang and Zhang (2015), Colliard and Demange (2014), Farboodi, Jarosch, and Shimer (2015), Hugonnier, Lester, and Weill (2014), Neklyudov (2015), and Shen, Wei, and Yan (2015) develop hybrid models, which are at the intersection of the search and the network literatures. The special case of my model with a homogeneous search intensity can be considered an extension of Hugonnier et al. (2014) with risk-averse investors and unrestricted asset holdings. They show that investors with average exogenous valuations specialize as intermediaries. In my setup with unrestricted holdings, investors with the "correct" amount of assets become intermediaries rather than the ones who have the average exogenous valuation. In other words, in my setup, intermediaries might be "low valuation-low holding," "average valuation-average holding," or "high valuation-high holding" investors. To my knowledge, in the literature, there are only two other papers with heterogeneity in search intensity: Neklyudov (2015) and Farboodi, Jarosch, and Shimer (2015). Both restrict the asset positions so that they lie in $\{0, 1\}$. Relative to these models, an important additional insight of my model is that fast investors can differentiate themselves from slow investors by offering more attractive trade quantities to their counterparties. In this way, they can charge a speed premium, and earn higher markups depending on the level of frictions. In the $\{0, 1\}$ models, fast investors typically earn lower markups because of the lower variability of their reservation values.

The remainder of the paper is organized as follows: Section 2 describes the model. Section 3 studies the equilibrium of the model, while Section 4 assesses the empirical implications of the endogenous asset positions in OTC markets given by the equilibrium. Section 5 is the conclusion.

2 Environment

Time is continuous and runs forever. I fix a probability space $(\Omega, \mathcal{F}, \Pr)$ and a filtration $\{\mathcal{F}_t, t \geq 0\}$ of sub-σ-algebras satisying the usual conditions (see Protter, 2004). There is a continuum of investors with a total measure normalized to 1. There is one long-lived asset in fixed supply denoted by A. This asset is traded over the counter, and pays an expected dividend flow denoted by m_D. There is also a perishable good, called the *numéraire*, which all investors produce and consume.

2.1 Preferences

I borrow the specification of preferences and trading motives from Duffie et al. (2007). Investors' level of risk aversion and time preference rate are denoted by γ and r respectively. The instantaneous utility function of an investor is $u(\rho, a) + c$, where

$$u(\rho, a) \equiv a m_D - \frac{1}{2} r \gamma \left(a^2 \sigma_D^2 + 2 \rho a \sigma_D \sigma_\eta \right) \tag{1}$$

is the instantaneous mean-variance benefit to the investor from holding $a \in \mathbb{R}$ units of the asset when of type $\rho \in [-1, +1]$, and $c \in \mathbb{R}$ denotes the net consumption of the numéraire good. An investor's net consumption becomes negative when she produces the numéraire to make side payments.

This utility specification is interpreted in terms of risk aversion. Since the parameter m_D is an expected rather than actual dividend flow, this cash flow needs to be adjusted for risk. The term $a^2\sigma_D^2$ represents the instantaneous variance of the asset payoff where σ_D is the volatility of the asset payoff. The term $2\rho a \sigma_D \sigma_\eta$ captures the instantaneous covariance between the asset payoff and some background risk with volatility σ_η. Therefore, the investor's type ρ captures the instantaneous correlation between the asset payoff and the background risk. In Appendix A, I derive this mean-variance utility specification from first principles.[8] I leave the microfoundation of this specification to the Appendix because the reduced-form imparts the main intuitions without the burden of derivations.

Importantly, the correlation between the asset payoff and the background risk is heterogeneous across investors, creating the gains from trade. In the context of different markets, this heterogeneity can be interpreted in different ways such as hedging demands or liquidity needs. In the case of a credit derivatives market, for example, the correlation captures the exposure to credit risk. If a bank's exposure to the credit risk of a certain bond or loan is high, the correlation between the bank's income and the payoff of the derivative written on that specific bond or loan will be negative, implying that the derivative provides hedging to the bank. Therefore, that bank will have a high valuation for the derivative. Another bank with a short position in the bond will have a positive correlation and, consequently, a low valuation for the derivative.

[8] I assume that investors have CARA preferences over the numéraire good, and they can invest in a riskless asset traded in a Walrasian market, and in a risky asset traded over the counter. Moreover, the investor receives a random income whose correlation with the payoff of risky asset is ρ. These assumptions give rise to my reduced-form specification, up to a suitable first-order approximation. See Duffie et al. (2007), Vayanos and Weill (2008), and Gârleanu (2009) for a similar derivation.

I assume that each investor's type itself is stochastic. Namely, an investor receives idiosyncratic correlation shocks at Poisson arrival times with intensity $\alpha > 0$. Arrival of these shocks is independent from other stochastic processes and across investors. For simplicity, I assume that types are not persistent, and upon the arrival of an idiosyncratic shock, the investor's new type is drawn according to the cdf F on $[-1, +1]$.

2.2 Trade

All trades are fully bilateral. I assume that investors with different search efficiencies co-exist in a sense that will now be described.

Following Weill (2008), I assume that investor i is endowed with a measure λ_i of "trading specialists," who search for other investors' trading specialists for trade opportunities. The measure of an investor's trading specialists determine how efficiently she searches. A given specialist finds a counterparty with an intensity $\mu > 0$, reflecting the overall search efficiency of the market. Therefore, investor i finds a counterparty at total instantaneous rate $\mu\lambda_i$. Conditional on contact, the counterparty is chosen randomly from the pool of all trading specialists.

The cross-sectional distribution of the measure of trading specialists is given by cdf $\Psi(\lambda)$ on $[0, 1]$.[9] The parameter λ is distributed independently from the correlation type ρ in the cross-section, and from all the stochastic processes in the model. Each contact between investor (ρ, a, λ) and investor (ρ', a', λ') is followed by a symmetric Nash bargaining game over quantity $q[(\rho, a, \lambda), (\rho', a', \lambda')]$ and unit price $P[(\rho, a, \lambda), (\rho', a', \lambda')]$. The number of assets, the investor (ρ, a, λ) purchases, is denoted by $q[(\rho, a, \lambda), (\rho', a', \lambda')]$. Thus, she will become an investor of type $(\rho, a + q[(\rho, a, \lambda), (\rho', a', \lambda')], \lambda)$ after this trade, while her counterparty will become type $(\rho', a' - q[(\rho, a, \lambda), (\rho', a', \lambda')], \lambda')$. The per unit price, the investor (ρ, a, λ)

[9] Because scaling μ and all λs up and down, respectively, by the same factor has no effect, I normalize the upper bound of the support to 1.

will pay, is denoted by $P\left[(\rho, a, \lambda), (\rho', a', \lambda')\right]$.

3 Equilibrium

In this section, I define a stationary equilibrium for this economy. Then, as a benchmark case, I solve the Walrasian counterpart of this economy. Finally, I characterize the stationary decentralized market equilibrium.

3.1 Definition

First, I will define the investors' value functions, taking as given the equilibrium joint distribution of investor types, asset holdings, and the measure of trading specialists. Then, I will write down the conditions that the equilibrium distribution satisfies.

3.1.1 Investors

Let $J(\rho, a, \lambda)$ be the maximum attainable utility of an investor of type (ρ, a, λ). In steady state, the Bellman principle implies that the growth rate of any investor's continuation utility must be the discount rate r (see Duffie, 2012a). Thus, it satisfies

$$\begin{aligned} rJ(\rho, a, \lambda) &= u(\rho, a) + \alpha \int_{-1}^{1} [J(\rho', a, \lambda) - J(\rho, a, \lambda)] dF(\rho') \\ &+ \int_{0}^{1} \int_{-\infty}^{\infty} \int_{-1}^{1} 2\mu\lambda \frac{\lambda'}{\Lambda} \left\{ J(\rho, a + q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right], \lambda) - J(\rho, a, \lambda) \right. \\ &\left. - q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right] P\left[(\rho, a, \lambda), (\rho', a', \lambda')\right] \right\} \Phi(d\rho', da', d\lambda'), \quad (2) \end{aligned}$$

where

$$\begin{aligned}&\{q\left[(\rho,a,\lambda),(\rho',a',\lambda')\right],P\left[(\rho,a,\lambda),(\rho',a',\lambda')\right]\}\\&=\arg\max_{q,P}\left[J(\rho,a+q,\lambda)-J(\rho,a,\lambda)-Pq\right]^{\frac{1}{2}}\left[J(\rho',a'-q,\lambda')-J(\rho',a',\lambda')+Pq\right]^{\frac{1}{2}},\qquad(3)\end{aligned}$$

s.t.

$$J(\rho,a+q,\lambda)-J(\rho,a,\lambda)-Pq\geq 0,$$
$$J(\rho',a'-q,\lambda')-J(\rho',a',\lambda')+Pq\geq 0.$$

The first term on the RHS of the equation (2) is the investor's utility flow; the second term is the expected change in the investor's continuation utility, conditional on switching types, which occurs with Poisson intensity α; and the third term is the expected change in the continuation utility, conditional on trade, which occurs with Poisson intensity $2\mu\lambda$. The potential counterparty is drawn randomly from the population, with the likelihood, $\frac{\lambda'}{\Lambda}$, that is proportional to her measure of trading specialists, where $\Lambda=\int_0^1\lambda' d\Psi\left(\lambda'\right)$.[10] The joint cdf of the stationary distribution of types, asset holdings, and search intensities is $\Phi(\rho',a',\lambda')$. Terms of trade, $q\left[(\rho,a,\lambda),(\rho',a',\lambda')\right]$ and $P\left[(\rho,a,\lambda),(\rho',a',\lambda')\right]$, maximize the symmetric Nash product (3) subject to the usual individual rationality constraints.

3.1.2 Market clearing and the distribution of investors' states

Let $\Phi(\rho^*,a^*,\lambda^*)$ denote the joint cumulative distribution of correlations, asset holdings, and the measure of specialists in the stationary equilibrium. Since $\Phi(\rho^*,a^*,\lambda^*)$ is a joint cdf, it

[10]The total matching rate is $2\mu\lambda$ because the investor finds a counterparty at rate $\int_0^1\mu\lambda\frac{\lambda'}{\Lambda}d\Psi\left(\lambda'\right)$, and another investor finds her at rate $\int_0^1\mu\lambda'\frac{\lambda}{\Lambda}d\Psi\left(\lambda'\right)$. This matching function is a variant of the CRS matching function of Shimer and Smith (2001).

should satisfy

$$\int_{0}^{1}\int_{-\infty}^{\infty}\int_{-1}^{1}\Phi(d\rho^*, da^*, d\lambda^*) = 1. \tag{4}$$

The clearing of the market for the asset requires that

$$\int_{0}^{1}\int_{-\infty}^{\infty}\int_{-1}^{1}a^*\Phi(d\rho^*, da^*, d\lambda^*) = A. \tag{5}$$

Since the heterogeneity in search intensity is *ex ante*, I impose

$$\int_{0}^{\lambda^*}\int_{-\infty}^{\infty}\int_{-1}^{1}\Phi(d\rho, da, d\lambda) = \Psi\left(\lambda^*\right) \tag{6}$$

for all $\lambda^* \in \text{supp}(\Psi)$ to ensure that the equilibrium distribution is consistent with the cross-sectional distribution of λs.

Finally, the conditions for stationarity are

$$\begin{aligned} &-\alpha\Phi(\rho^*, a^*, \lambda^*)(1 - F(\rho^*)) + \alpha\int_{0}^{\lambda^*}\int_{-\infty}^{a^*}\int_{\rho^*}^{1}\Phi(d\rho, da, d\lambda)F(\rho^*) \\ &-\int_{0}^{\lambda^*}\int_{-\infty}^{a^*}\int_{-1}^{\rho^*}\left[\int_{0}^{1}\int_{-\infty}^{\infty}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}\mathbb{I}_{\{q[(\rho,a,\lambda),(\rho',a',\lambda')]\geq a^*-a\}}\Phi(d\rho', da', d\lambda)\right]\Phi(d\rho, da, d\lambda) \\ &+\int_{0}^{\lambda^*}\int_{a^*}^{\infty}\int_{-1}^{\rho^*}\left[\int_{0}^{1}\int_{-\infty}^{\infty}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}\mathbb{I}_{\{q[(\rho,a,\lambda),(\rho',a',\lambda')]<a^*-a\}}\Phi(d\rho', da', d\lambda)\right]\Phi(d\rho, da, d\lambda) = 0 \end{aligned} \tag{7}$$

for all $(\rho^*, a^*, \lambda^*) \in \text{supp}(\Phi)$.

The first term of the first line is the outflow due to idiosyncratic shocks. Investors who belong to $\Phi(\rho^*, a^*, \lambda^*)$ receive correlation shocks at rate α, and they leave $\Phi(\rho^*, a^*, \lambda^*)$ with probability $1 - F(\rho^*)$, i.e., if their new type is higher than ρ^*. Similarly, the second term of the first line is the inflow due to idiosyncratic shocks. Investors, who do not belong to

$\Phi(\rho^*, a^*, \lambda^*)$ but have an asset holding less than a^* and a total measure of specialists less than λ^*, receive correlation shocks at rate α, and they enter $\Phi(\rho^*, a^*, \lambda^*)$ with probability $F(\rho^*)$, i.e., if their new type is less than ρ^*.

The second line represents the outflow due to trade. Conditional on a contact, investors, who belong to $\Phi(\rho^*, a^*, \lambda^*)$, leave $\Phi(\rho^*, a^*, \lambda^*)$ if they buy a sufficiently high number of assets, i.e., if they buy at least $a^* - a$ units where a is the number of assets before trade. Similarly, the third line represents the inflow due to trade. Investors, who do not belong to $\Phi(\rho^*, a^*, \lambda^*)$ but have a correlation less than ρ^* and a total measure of specialists less than λ^* enter $\Phi(\rho^*, a^*)$ if they sell a sufficiently high number of assets, i.e., if they sell at least $a - a^*$ units, where a is the number of assets before trade. Note that selling at least $a - a^*$ units is equivalent to buying at most $a^* - a$ units, and, hence, I write $q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right] < a^* - a$ inside the indicator function.

A stationary equilibrium is defined as follows:

Definition 1 *A stationary equilibrium is (i) a pricing function* $P\left[(\rho, a, \lambda), (\rho', a', \lambda')\right]$, *(ii) a trade size function* $q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right]$, *(iii) a function* $J(\rho, a, \lambda)$ *for continuation utilities, and (iv) a joint distribution* $\Phi(\rho, a, \lambda)$ *of types, asset holdings, and the measure of specialists, such that*

- *Steady-state: Given ii), iv) solves the system (4)-(7).*
- *Optimality: Given i), ii), and iv), iii) solves the investor's problem (2) subject to (3).*
- *Nash bargaining: Given iii), i) and ii) satisfy (3).*

3.2 The Walrasian benchmark

I solve the stationary equilibrium of a continuous frictionless Walrasian market as a benchmark. Then, I use the outcome of this benchmark to better understand the effect of trading frictions on market outcomes. Since, in this market, every investor can trade instantly, there is one market-clearing price and all investors with the same correlation type hold the same number of assets. The flow Bellman equation of investors in this Walrasian market is

$$rJ^W(\rho,a) = u(\rho,a) + \alpha\int_{-1}^{1} \max_{a'} \left\{J^W(\rho',a') - J^W(\rho,a) - P^W(a'-a)\right\} dF(\rho'),$$

where P^W is the market-clearing price. The first term is the investor's utility flow. The second term is the expected change in the investor's continuation utility, conditional on switching types, which occurs with Poisson intensity α. Since investors have continuous access to the market, they rebalance their holding as soon as they receive an idiosyncratic shock. The FOC for the asset position and the envelope condition[11] are

$$J_2^W(\rho',a') = P^W$$

and

$$rJ_2^W(\rho,a) = u_2(\rho,a) + \alpha\left(-J_2^W(\rho,a) + P^W\right),$$

where $u_2(.,.)$ represents the partial derivative with respect to the second argument. Combining these two conditions, I get the optimal demand of the investor with ρ:

$$a^W(\rho;P) = \frac{1}{\gamma\sigma_D^2}\left(\frac{m_D}{r} - P^W\right) - \frac{\sigma_\eta}{\sigma_D}\rho.$$

[11] To write down these conditions, I assume that $J^W(\rho,.)$ is stricly concave and continuosly differentiable. This assumption is verified *ex post*.

The market-clearing condition

$$\int_{-1}^{1} a^W(\rho; P) dF(\rho) = A$$

implies that the equilibrium objects are:

$$a^W(\rho) = A - \frac{\sigma_\eta}{\sigma_D} (\rho - \overline{\rho})$$

for all $\rho \in supp(F)$; and

$$P^W = \frac{u_2(\overline{\rho}, A)}{r} = \frac{m_D}{r} - \gamma\sigma_D^2 A - \gamma\sigma_D\sigma_\eta\overline{\rho},$$

where

$$\overline{\rho} \equiv \int_{-1}^{1} \rho' dF(\rho').$$

The implication of the equilibrium is intuitive: The equilibrium holding is a decreasing function of correlation ρ. As ρ increases, the hedging benefit of the asset decreases and investors hold less of it. The investor with the average correlation holds the per capita supply. The coefficient of the current correlation in the optimal holding is $\frac{\sigma_\eta}{\sigma_D}$. The volatility of the background risk, σ_η, has a positive impact on the dispersion of investors' holdings because they have a higher incentive to hold or stay away from the asset when their background is more volatile. On the other hand, the volatility of the asset payoff, σ_D, has a negative impact on the dispersion of investors' holdings because the importance of the cost of risk-bearing relative to the hedging demand rises when the asset payoff is more volatile. Thus, investors' positions become closer to each other as required by efficient risk-sharing.

The instantaneous trading volume in the Walrasian market is

$$\mathbb{V}^{W} = \alpha \int_{-1}^{1}\int_{-1}^{1} \left|a^{W}(\rho') - a^{W}(\rho)\right| dF(\rho)\, dF(\rho') = \alpha \frac{\sigma_\eta}{\sigma_D} \int_{-1}^{1}\int_{-1}^{1} |\rho' - \rho|\, dF(\rho)\, dF(\rho').$$

This is basically the multiplication of the flow of investors who receive idiosyncratic shock, α, and the change in the optimal holding of those investors. When I characterize the OTC market equilibrium, I will show that the Walrasian market outcomes differ markedly from the OTC outcomes. As a preview, in the Walrasian equilibrium, (i) there is no price dispersion, (ii) no one provides intermediation (apart from the Walrasian auctioneer), and, therefore, (iii) net and gross trade volume coincide.

Finally, I calculate the sum of all investors' continuation utilities as a measure of welfare, following Gârleanu (2009):

$$\mathbb{W}^{W} = \frac{m_D}{r} A - \frac{\gamma \sigma_D^2}{2} A^2 - \gamma \sigma_D \sigma_\eta \bar{\rho} A + \frac{\gamma \sigma_\eta^2}{2} var[\rho].$$

The last term of the welfare exclusively captures the hedging benefit from being able to access the centralized market instantly following an idiosyncratic shock. The frictions of the OTC market will affect the welfare through this term.

3.3 Characterization

3.3.1 Individual trades

Terms of individual trades, $q[(\rho, a, \lambda), (\rho', a', \lambda')]$ and $P[(\rho, a, \lambda), (\rho', a', \lambda')]$, are determined by a Nash bargaining game with the solution given by the optimization problem (3). I guess and verify that $J(\rho, ., \lambda)$ is continuously differentiable and strictly concave for all ρ and λ. This allows me to set up the Lagrangian of this problem, and find the first-order necessary and sufficient conditions (see Theorem M.K.2., p. 959, and Theorem M.K.3., p. 961, in

Mas-Colell, Whinston & Green, 1995) for optimality by differentiating the Lagrangian. The trade size, $q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right]$, solves

$$J_2(\rho, a+q, \lambda) = J_2(\rho', a'-q, \lambda'), \tag{8}$$

where J_2 represents the partial derivative with respect to the second argument. Notice that the quantity which solves the equation (8) is also the maximizer of the total trade surplus, i.e.,

$$q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right] = \arg\max_q \; J(\rho, a+q, \lambda) - J(\rho, a, \lambda) + J(\rho', a'-q, \lambda') - J(\rho', a', \lambda').$$

The continuous differentiability and strict concavity of $J(\rho, ., \lambda)$ guarantees the existence and uniqueness of $q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right]$. Then, the transaction price, $P\left[(\rho, a, \lambda), (\rho', a', \lambda')\right]$, is determined such that the total trade surplus is split equally between the parties:

$$P = \frac{J(\rho, a+q, \lambda) - J(\rho, a, \lambda) - (J(\rho', a'-q, \lambda') - J(\rho', a', \lambda'))}{2q} \tag{9}$$

if $J_2(\rho, a, \lambda) \neq J_2(\rho', a', \lambda')$; and $P = J_2(\rho, a, \lambda)$ if $J_2(\rho, a, \lambda) = J_2(\rho', a', \lambda')$. Substituting the trade quantity and price into (2), I get

$$\begin{aligned} rJ(\rho, a, \lambda) &= u(\rho, a) + \alpha \int_{-1}^{1} [J(\rho', a, \lambda) - J(\rho, a, \lambda)] dF(\rho') \\ &+ \int_0^1 \int_{-\infty}^{\infty} \int_{-1}^{1} 2\mu\lambda \frac{\lambda'}{\Lambda} \frac{1}{2} \left[\max_q \; \{J(\rho, a+q, \lambda) - J(\rho, a, \lambda) \right. \\ &\left. + J(\rho', a'-q, \lambda') - J(\rho', a', \lambda')\} \right] \Phi(d\rho', da', d\lambda'). \end{aligned} \tag{10}$$

In order to solve for $J(\rho, a, \lambda)$, I follow a guess-and-verify approach. The complete solution is given in the Appendix. In the models with $\{0,1\}$ holding, investors' trading behavior is determined by their reservation value, which is the difference between the value of holding the asset and that of not holding the asset. The counterpart of the reservation value in my model with unrestricted holdings is the marginal continuation utility or the marginal valuation in short. To find the marginal valuation, I differentiate the equation (10) with respect to a, applying the envelope theorem:

$$rJ_2(\rho, a, \lambda) = u_2(\rho, a) + \alpha \int_{-1}^{1} [J_2(\rho', a, \lambda) - J_2(\rho, a, \lambda)] dF(\rho')$$
$$+ \int_{0}^{1} \int_{-\infty}^{\infty} \int_{-1}^{1} \mu\lambda \frac{\lambda'}{\Lambda} \{J_2(\rho, a + q[(\rho, a, \lambda), (\rho', a', \lambda')], \lambda) - J_2(\rho, a, \lambda)\} \Phi(d\rho', da', d\lambda'), \quad (11)$$

where

$$u_2(\rho, a) = m_D - r\gamma\sigma_D^2 a - r\gamma\sigma_D\sigma_\eta\rho.$$

Since the utility function is quadratic, the marginal utility flow is linear. The equation (11) is basically a flow Bellman equation that has a linear return function with a slope coefficient independent of ρ. Therefore, the solution $J_2(\rho, a, \lambda)$ is linear in a if and only if $q[(\rho, a, \lambda), (\rho', a', \lambda')]$ is linear in a. Conjecturing that $q[(\rho, a, \lambda), (\rho', a', \lambda')]$ is linear in a, and that the slope coefficient of a in the marginal valuation is $-\frac{r\gamma\sigma_D^2}{\widetilde{r}(\lambda)}$ for $\widetilde{r}(\lambda) > 0$,[12] the FOC (8) implies that

$$J_2(\rho, a + q[(\rho, a, \lambda), (\rho', a', \lambda')], \lambda) = \frac{\widetilde{r}(\lambda) J_2(\rho, a, \lambda) + \widetilde{r}(\lambda') J_2(\rho', a', \lambda')}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')}, \quad (12)$$

i.e., the post-trade marginal valuation of both investors is equal to the weighted average of their initial marginal valuations with the weights being the reciprocal of the slope coefficient

[12] These conjectures are verified in the proof of Theorem 1.

of a in the marginal valuation. Note that the post-trade marginal valuation will be equal to the midpoint of the investors' initial marginal valuations if they are endowed with the same measure of specialists.

In principle, optimal trading rules, interacting in complex ways with the equilibrium distribution, make a fully bilateral trade model with unrestricted holdings difficult to solve. So far, the literature has side-stepped this difficulty by considering models with value functions that can be characterized before solving for the endogenous distribution. This is not the case in my model. As can be seen from (11) and (12), search intensity interacts with correlation and asset holding in the Bellman equation for the marginal valuation. The problem becomes relatively easy because (i) correlation and asset holding are in separate terms in the marginal utility, and (ii) the distribution of correlations and the distribution of search intensities are independent. Thanks to these assumptions, search intensity interacts only with asset holding. As a result, I need to solve for the average asset holding conditional on λ. This creates a fixed point problem which requires solving a linear system for the average asset holding conditional on λ and the average marginal valuation conditional on λ. The equations of the system come from optimality conditions, steady-state conditions and the market clearing. Its unique solution implies that the average asset holding conditional on λ is the supply A, which is independent of λ, i.e., the primary effect of heterogeneity in λ will be to affect the variance and the higher order moments of the distribution. This allows me to obtain the following theorem from the equation (11):

Theorem 1 *In any stationary equilibrium, investors' marginal valuations satisfy*

$$J_2(\rho, a, \lambda) = \frac{m_D - r\gamma\sigma_D^2 a - r\gamma\sigma_D\sigma_\eta \frac{\widetilde{r}(\lambda)\rho + \alpha\overline{\rho}}{\widetilde{r}(\lambda) + \alpha} + \left(\widetilde{r}(\lambda) - r\right)\frac{u_2(\overline{\rho}, A)}{r}}{\widetilde{r}(\lambda)}, \tag{13}$$

where

$$\widetilde{r}(\lambda) = r + \int_0^1 \mu\lambda \frac{\lambda'}{\Lambda} \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')} d\Psi(\lambda'). \tag{14}$$

And, the average marginal valuation of the market is

$$\int_0^1 \int_{-\infty}^{\infty} \int_{-1}^{1} J_2(\rho, a, \lambda)\, \Phi(d\rho, da, d\lambda) = \frac{u_2(\overline{\rho}, A)}{r}. \tag{15}$$

Equation (13) shows that an investor's marginal valuation equals the combination of her current expected marginal utility flow until the next trade opportunity (the first three terms) and the expected contribution of the market to her post-trade marginal valuation (the last term). In this characterization, $\widetilde{r}(\lambda)$ has a natural interpretation as the *effective discount rate* of an investor with λ as it is the actual rate at which the investor discounts the current marginal utility flow associated with her current asset holding. Therefore, the effective discount rate is an important determinant of the sensitivity of an investor's marginal valuation to her asset holding. In addition, an alternative environment where investors have access to a centralized market at Poisson arrival times with intensity $\widetilde{r}(\lambda) - r$ would lead to the same marginal valuation in (13). After every trade, the trading investor's marginal valuation would be equal to the average marginal valuation of the market. In this sense, the effective discount rate (14) can be understood as the sum of discount rate, r, and the (effective) transition rate to the post-trade state.

Although the effective discount rates are not available in closed form for an arbitrary distribution of the measure of specialists, most of the important qualitative implications of heterogeneity in the measure of specialists come from the properties stated in Lemma 1. In particular, it states that the effective discount rate is an increasing function of λ. An important implication of this combined with (13) is that the marginal valuation of investors with high λ is closer to the average marginal valuation of the market, controlling for asset holding and hedging need. Therefore, investors with high λ become the natural counterparty for investors with high marginal valuations and those with low marginal valuations. They buy the assets from investors with low marginal valuations and sell to investors with high marginal valuations, thus, become endogenous "middlemen".

Lemma 1 *Suppose the support of the distribution, Ψ, is finite. Then, an effective discount rate function, $\widetilde{r}(\lambda)$, which is consistent with the optimality of the investors' problem, exists, is unique, strictly increasing and strictly concave, and satisfies*

$$\int_0^1 \widetilde{r}(\lambda)\, d\Psi(\lambda) = r + \frac{\mu\Lambda}{2},$$

where

$$\Lambda \equiv \int_0^1 \lambda' d\Psi(\lambda').$$

The functional equation (14) shows two key properties of the effective discount rate: being increasing and concave. On the one hand, the measure of trading specialists has a direct linear positive impact on the effective discount rate. If an investor is able to find counterparties very often, she does not expect to spend much time with her current holding, and her marginal valuation should depend less on her current marginal utility flow. Hence, she should discount her current marginal utility at a higher rate. This makes the effective discount rate an increasing function. On the other hand, equation (12) shows that the post-trade marginal valuation is closer to the initial marginal valuation of the party with higher effective discount rate. Because of this, a high search intensity dampens the effect of trade on post-trade marginal valuation. Thus, an indirect negative impact of λ on the effective discount rate arises. Consequently, the effective discount rate turns out to be an increasing but concave function of λ.

Again, using the fact that $J(\rho, a, \lambda)$ is quadratic in a, an exact second-order Taylor expansion shows that:

$$J(\rho, a+q, \lambda) - J(\rho, a, \lambda) = J_2(\rho, a+q, \lambda)q + \frac{r\gamma\sigma_D^2}{2\widetilde{r}(\lambda)}q^2.$$

Next, Equation (9) implies

$$\begin{aligned} P\left[(\rho, a, \lambda), (\rho', a', \lambda')\right] &= J_2(\rho, a + q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right], \lambda) \\ &\quad + \frac{r\gamma\sigma_D^2}{4} q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right]\left(\frac{1}{\widetilde{r}(\lambda)} - \frac{1}{\widetilde{r}(\lambda')}\right). \end{aligned} \tag{16}$$

i.e., the transaction price is given by the post-trade marginal valuation plus an adjustment term. I call the adjustment term the "speed premium" because it always benefits the investor who is able to find counterparties faster. Note that the transaction price will be equal to the post-trade marginal valuation if the trading parties have the same speed. This formula for the price explains the main mechanism behind the relation between λ and intermediation markups. Due to the first term, investors with high λ tend to earn lower markups since they have stable marginal valuations that do not fluctuate much in response to changes in asset holding and hedging need. On the other hand, they earn a premium that is increasing in trade size. Thus, in equilibrium, if trade sizes are large enough, the second term dominates and fast investors earn higher markups. If trade sizes are small enough, the first term dominates and fast investors earn lower markups. Consequently, my model rationalizes both *the centrality premium* and *the centrality discount* in intermediation markups, which are empirically documented in distinct works.[13]

In equilibrium, investors who trade in high quantities are the ones who have received an idiosyncratic shock recently. After the arrival of an idiosyncratic shock, the investor's first few trades mostly reflect her effort to get close to her new ideal asset position. During this

[13] Li and Schürhoff (2012) and Bech and Atalay (2010) find that central dealers earn higher markups in the municipal bond market and the fed funds market, respectively. Hollifield et al. (2014) find that central dealers earn lower markups in the market for asset-backed securities.

period, she trades in higher quantities than she does when she is close to her ideal position. Hence, if investors spend too much time following an idiosyncratic shock until they become close to their new ideal position, fast investors have the opportunity to earn substantial speed premia. Given a distribution of search intensities and a distribution of correlations, this is determined by the aggregate level of frictions in the market. More specifically, if the intensity of idiosyncratic shocks, α, is high, and the aggregate search efficiency, μ, is low, this becomes the case. Therefore, in markets with a high level of frictions, the speed premium dominates and we observe a centrality premium in intermediation markups. In markets with a low level of frictions, we observe a centrality discount in intermediation markups.

The next proposition shows analytically how terms of trade depend on investors' current state.

Proposition 1 *Let*

$$\theta(\rho, a, \lambda) = A - a + \frac{\sigma_\eta}{\sigma_D} \frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda) + \alpha} (\overline{\rho} - \rho)$$

denote the effective type of the investor with (ρ, a, λ)*. In any stationary equilibrium, investors' marginal valuations, individual trade sizes, and transaction prices are given by:*

$$J_2(\rho, a, \lambda) = \frac{u_2(\overline{\rho}, A)}{r} + \frac{r\gamma\sigma_D^2}{\widetilde{r}(\lambda)} \theta(\rho, a, \lambda), \tag{17}$$

$$q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right] = \frac{\frac{1}{\widetilde{r}(\lambda)}\theta(\rho, a, \lambda) - \frac{1}{\widetilde{r}(\lambda')}\theta(\rho', a', \lambda')}{\frac{1}{\widetilde{r}(\lambda)} + \frac{1}{\widetilde{r}(\lambda')}} \tag{18}$$

and

$$P\left[(\rho, a, \lambda), (\rho', a', \lambda')\right] = \frac{u_2(\overline{\rho}, A)}{r} + r\gamma\sigma_D^2 \frac{\frac{3\widetilde{r}(\lambda)+\widetilde{r}(\lambda')}{4\widetilde{r}(\lambda)}\theta(\rho, a, \lambda) + \frac{\widetilde{r}(\lambda)+3\widetilde{r}(\lambda')}{4\widetilde{r}(\lambda')}\theta(\rho', a', \lambda')}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')}. \tag{19}$$

If there were no heterogeneity in ρ or in λ, the quantity traded in a bilateral meeting would depend only on pre-trade asset positions as in Afonso and Lagos (2015). In this sense, my model generalizes the trading rule of Afonso and Lagos (2015) by showing that, in my more general model, it depends also on preference parameters (r, σ_η, σ_D and α) and search efficiency parameters (μ, λ, λ'). This effect of the preference parameters on trading rules is a key channel through which changes in the OTC market frictions affect trading volume, price dispersion, and welfare, as I will show in Section 4 when I discuss the empirical implications of the model.

The effective type of Proposition 1 is a sufficient statistic for the effect of an investor's current state on her trading behavior. Indeed, the effective type of an investor is her ideal trade quantity stemming from optimal hedging behavior. Given that investors are trying to equalize their marginal valuations by correcting their holdings, θ represents the ideal trade quantity. Investors would be able to trade in these quantities if their counterparties had a constant marginal valuation of $\frac{u_2(\overline{\rho},A)}{r}$, i.e., $\theta(\rho, a, \lambda)$ satisfies

$$J_2(\rho, a+\theta, \lambda) = \frac{u_2(\overline{\rho}, A)}{r},$$

where $\frac{u_2(\overline{\rho},A)}{r}$ is the average marginal valuation of the market. If the effective type is 0, the investor's marginal valuation is equal to the average marginal valuation of the market. If she has a negative effective type, she has a lower than average marginal valuation of the market, and vice-versa. In a bilateral match between investors (ρ, a, λ) and (ρ', a', λ'), ideally the first party would want to buy $\theta(\rho, a, \lambda)$ units, and the second party would want to sell $-\theta(\rho', a', \lambda')$ units of the asset. Thus, the realized trade quantity (18) is a linear combination of the parties' ideal trade quantities with weights being the reciprocal of their effective discount rates. This is an important result because of its implications for the supply of liquidity services. Because the effective discount rate is an increasing function, the equation (18) reveals that the trade quantity reflects the trading need of the slower counterparty to a greater extent. In this sense, fast investors provide immediacy by trading according to their counterparties' needs. For an investor with a very high λ, the weight of her ideal trade quantity in the bilateral trade quantity is very small, so the disturbance her hedging need creates for her counterparty is very small. Her counterparty is able to buy from or sell to her in almost exactly the ideal amount. A speed premium in the price arises because of this asymmetry in how the trade quantity reflects the trading need of the counterparties. Having high λ increases the importance of the option value of search and decreses the importance of the current utility flow from holding the asset. Therefore, fast investors put less weight on their asset positions and more weight on their cash earnings when bargaining with a counterparty. Each bilateral negotiation results in a trade size that is more in line with the slower counterparty's hedging need and a trade price that contains a premium benefitting the faster counterparty. An investor can achieve the average marginal valuation by trading with the right counterparty (or the right sequence of counterparties). The key observation here is that if she trades with a fast counterparty, she will achieve the average marginal valuation relatively quickly. The trade-off an investor faces is between the fast correction of the asset position and paying a low price. That is how the speed premium arises optimally. Figure 1 graphically presents an example of how trade quantity and price arise as the result of a bilateral negotiation between two investors with different λs.

In Figure 1, each line represents the marginal valuation as a function of asset holding given a certain level of correlation. The steeper line represents the marginal valuation of a slow investor while the flatter line represents the marginal valuation of a fast investor. This is the direct result of Equation (13). Since the effective discount rate is increasing in λ, the slope of the marginal valuation line is lower for investors with high λ. Suppose that two blue dots on the graph represent the inital positions of two investors. If they make contact, the investor on top will be the buyer as she has a higher marginal valuation. Trade allows investors to move horizontally. Green lines with arrows show the quantity and the direction of the trade. The joint surplus of this trade is the sum of the shaded triangular areas. As can be seen, the impact of trade on the slow investor's marginal valuation is higher than the impact of trade on the fast investor's marginal valuation. As a result, the triangle for the fast investor (the seller) is smaller than the triangle for the slow investor (the buyer). If the price were equal to the post-trade marginal valuation, the slow investor's surplus would be bigger than the fast investor's surplus. That would violate the symmetric Nash bargaining. For this reason, the fast investor charges a speed premium to equalize the individual trade surpluses by extracting surplus from the slow investor. The other case, in which the fast investor is the buyer, is symmetric. In this case, the price becomes lower than the post-trade marginal valuation as a result of the speed premium the fast investor charges.

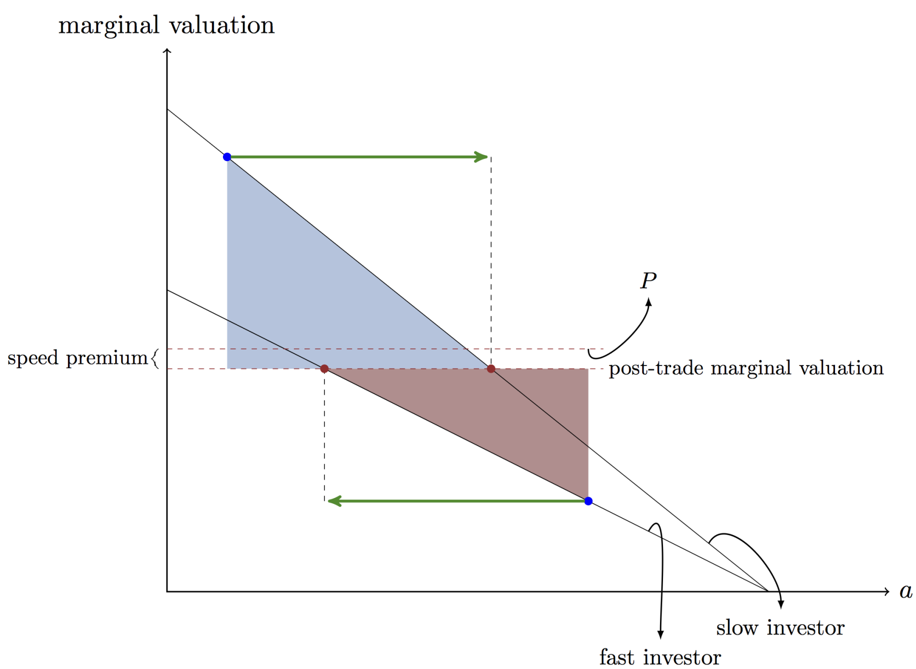

Figure 1. Sample trade between investors with different search intensities

An advantage of this setup is that the speed premium arises solely due to the differences in search intensity. In reality, fast investors might be more sophisticated and have higher bargaining power, and this might give rise to additional premia in prices. However, I show that the speed premium arises even when there is no asymmetry in terms of bargaining power.

3.3.2 The joint distribution of types, holdings, and search intensities

For simplicity, I assume that the distribution of correlations has a continuous support. In this case, the equilibrium conditional distributions of asset holdings have densities. This assumption is actually not necessary for the full characterization of the equilibrium distribution, but it simplifies the presentation of Proposition 2 as an intermediate step. Since I have an explicit expression for trade sizes, I can eliminate indicator functions in Equation (7). Writing the system of steady-state equations in terms of conditional pdfs $\phi_{\rho,\lambda}(a)$, I derive the following proposition:

Proposition 2 *In any stationary equilibrium, the conditional pdf $\phi_{\rho,\lambda}(a)$ of asset holdings satisfies the system*

$$\begin{aligned}(\alpha+2\mu\lambda)\,\phi_{\rho,\lambda}(a) = {} & \alpha\int_{-1}^{1}\phi_{\rho',\lambda}(a)\,dF(\rho') \\ & + \int_{0}^{1}\int_{-1}^{1}\int_{-\infty}^{\infty}2\mu\lambda\frac{\lambda'}{\Lambda}\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)\phi_{\rho,\lambda}(a') \\ \phi_{\rho',\lambda'}&\left(a\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)-a'+\overline{C}\left[(\rho,\lambda),(\rho',\lambda')\right]\right)da'dF(\rho')\,d\Psi(\lambda'), \quad (20)\end{aligned}$$

for all $(\rho,a,\lambda)\in supp(\Phi)$;

$$\int_{-\infty}^{\infty}\phi_{\rho,\lambda}(a)da = 1 \quad (21)$$

for all $\lambda\in supp(\Psi)$ and $\rho\in supp(F)$; and

$$\int_{0}^{1}\int_{-1}^{1}\int_{-\infty}^{\infty}a\phi_{\rho,\lambda}(a)dadF(\rho)d\Psi(\lambda) = A, \quad (22)$$

where

$$\overline{C}\left[\left(\rho,\lambda\right),\left(\rho',\lambda'\right)\right]\equiv\widetilde{r}\left(\lambda'\right)\frac{\sigma_{\eta}}{\sigma_{D}}\left(\frac{\rho-\overline{\rho}}{\widetilde{r}\left(\lambda\right)+\alpha}-\frac{\rho'-\overline{\rho}}{\widetilde{r}\left(\lambda'\right)+\alpha}\right)-\left[\frac{\widetilde{r}\left(\lambda'\right)}{\widetilde{r}\left(\lambda\right)}-1\right]A. \tag{23}$$

Equation (21) implies that $\phi_{\rho,\lambda}(a)$ is a pdf. Equation (22) is the market-clearing condition. Equation (20) has the usual steady-state interpretation. The first term represents the outflow due to idiosyncratic shocks and trade. The second and third terms represent the inflow due to idiosyncratic shocks and the inflow due to trade, respectively. The last term is an "adjusted" convolution (i.e., a convolution after an appropriate change of variable) since any investor of type (ρ, a', λ) can become one of type (ρ, a, λ) if she meets the right counterparty. The right counterparty in this context means an investor of type $\left(\rho', a\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)-a'+\overline{C}\left[\left(\rho,\lambda\right),\left(\rho',\lambda'\right)\right],\lambda'\right)$. Proposition 1 immediately implies that the post-trade type of the first investor will be (ρ, a, λ), and, hence, she will create inflow. Since the convolution term complicates the computation of the distribution function, I will make use of Fourier transform.[14] I follow the definition of Bracewell (2000) for the Fourier transform:

$$\widehat{g}(z)=\int_{-\infty}^{\infty}e^{-i2\pi xz}g\left(x\right)dx,$$

where $\widehat{g}(.)$ is the Fourier transform of the function $g\left(.\right)$.

Let $\widehat{\phi}_{\rho,\lambda}(.)$ be the Fourier transform of the equilibrium conditional pdf $\phi_{\rho,\lambda}(.)$. Then the Fourier transform of the equations (20)-(22) are, respectively:

[14] Following Duffie and Manso (2007); Duffie, Malamud, and Manso (2009), Duffie, Giroux, and Manso (2010), Andrei and Cujean (2014), and Cujean and Praz (2015) also made use of convolution for distributions in the context of search and matching models.

$$0 = -\left(\alpha + 2\mu\lambda\right)\widehat{\phi}_{\rho,\lambda}\left(z\right) + \alpha\int_{-1}^{1}\widehat{\phi}_{\rho',\lambda}\left(z\right)dF\left(\rho'\right) \tag{24}$$

$$+\int_{0}^{1}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}e^{i2\pi\overline{C}[(\rho,\lambda),(\rho',\lambda')]\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}}\widehat{\phi}_{\rho,\lambda}\left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\right)\widehat{\phi}_{\rho',\lambda'}\left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\right)dF\left(\rho'\right)d\Psi\left(\lambda'\right)$$

for all $\lambda \in \text{supp}(\Psi)$, $\rho \in \text{supp}(F)$ and for all $z \in \mathbb{R}$;

$$\widehat{\phi}_{\rho,\lambda}(0) = 1 \tag{25}$$

for all $\lambda \in \text{supp}(\Psi)$ and $\rho \in \text{supp}(F)$; and

$$\int_{0}^{1}\int_{-1}^{1}\widehat{\phi}'_{\rho,\lambda}(0)dF\left(\rho\right)d\Psi\left(\lambda\right) = -i2\pi A. \tag{26}$$

The system (24)-(26) cannot be solved in closed form. However, it facilitates the calculation of the moments which are derivatives of the transform, with respect to z, at $z = 0$. Thus, the system allows me to derive a recursive characterization of the moments of the equilibrium conditional distribution.

Proposition 3 *The following system characterizes all moments of the equilibrium conditional distributions of asset holdings*:

$$
\begin{aligned}
&(\alpha + 2\mu\lambda)\,\mathbb{E}_\phi\left[a^n \mid \rho, \lambda\right] = \alpha \mathbb{E}_\phi\left[a^n|\lambda\right] \\
&\quad + \sum_{j_1+j_2+j_3=n} \begin{pmatrix} n \\ j_1, j_2, j_3 \end{pmatrix} \mathbb{E}_\phi\left[a^{j_2} \mid \rho, \lambda\right] \Bigg\{ \sum_{k_1+k_2+k_3=j_1} \begin{pmatrix} j_1 \\ k_1, k_2, k_3 \end{pmatrix} \left(\frac{\sigma_\eta}{\sigma_D}\right)^{k_1+k_2} \\
&\qquad \left(-\frac{\rho}{\widetilde{r}(\lambda) + \alpha}\right)^{k_1} \Bigg[\int_0^1 2\mu\lambda \frac{\lambda'}{\Lambda} \left(\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')}\right)^n \left(\widetilde{r}(\lambda')\right)^{k_1+k_2} \\
&\qquad\qquad \left(\frac{1}{\widetilde{r}(\lambda') + \alpha}\right)^{k_2} \left(D(\lambda, \lambda')\right)^{k_3} \mathbb{E}_\phi\left[a^{j_3}\rho^{k_2} \mid \lambda'\right] d\Psi(\lambda') \Bigg] \Bigg\} \qquad (27)
\end{aligned}
$$

for all $\lambda \in supp(\Psi)$, $\rho \in supp(F)$ *and for all* $z \in \mathbb{R}$; *and*

$$
\mathbb{E}_\phi\left[a \mid \lambda\right] = A \qquad (28)
$$

for all $\lambda \in supp(\Psi)$; *where*

$$
D(\lambda, \lambda') \equiv \left(\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)} - 1\right) \left[A + \frac{\sigma_\eta}{\sigma_D} \frac{\widetilde{r}(\lambda')\,\widetilde{r}(\lambda)}{(\widetilde{r}(\lambda) + \alpha)(\widetilde{r}(\lambda') + \alpha)} \overline{\rho}\right]. \qquad (29)
$$

I use this characterization to analyze various dimensions of aggregate market liquidity, such as expected prices, average trade sizes, price dispersion, and welfare.

3.4 Constrained inefficiency

In this subsection, I investigate whether the fully decentralized market structure with unrestricted holdings is able to reallocate the assets efficienctly. I take the frictions as given and ask how a benevolent social planner would choose the quantity of assets transfered conditional on a contact. I define social welfare as the discounted sum of the utility flows of all investors,

$$\mathbb{W} = \int_0^\infty e^{-rt} \left\{ \int_0^1 \int_{-\infty}^\infty \int_{-1}^1 u(\rho, a) \Phi_t(d\rho, da, d\lambda) \right\} dt. \tag{30}$$

Any transfer of the numéraire good from one investor to another does not enter $\mathbb{W}$ because of quasi-linear preferences. The planner maximizes $\mathbb{W}$ with respect to the time path for the state variables, $\Phi_t(\rho, a, \lambda)$, and controls, $q_t^*[(\rho, a, \lambda), (\rho', a', \lambda')]$, subject to the laws of motion for these state variables and to the feasibility condition of asset reallocation,

$$q_t^*[(\rho, a, \lambda), (\rho', a', \lambda')] + q_t^*[(\rho', a', \lambda'), (\rho, a, \lambda)] = 0, \tag{31}$$

which also results in the imposition that the solution does not depend on the identities or "names" of investors. The planner's current-value Hamiltonian can be written as

$$\begin{aligned} L = & \int_0^1 \int_{-\infty}^\infty \int_{-1}^1 u(\rho, a) \Phi(d\rho, da, d\lambda) \\ & + \alpha \int_{-1}^1 \int_0^1 \int_{-\infty}^\infty \int_{-1}^1 (\vartheta(\rho', a, \lambda) - \vartheta(\rho, a, \lambda)) \Phi(d\rho, da, d\lambda) dF(\rho') \\ & + \int_0^1 \int_{-\infty}^\infty \int_{-1}^1 \int_0^1 \int_{-\infty}^\infty \int_{-1}^1 2\mu\lambda \frac{\lambda'}{\Lambda} \{\vartheta(\rho, a + q^*[(\rho, a, \lambda), (\rho', a', \lambda')], \lambda) - \vartheta(\rho, a, \lambda) \\ & + \vartheta(\rho', a' - q^*[(\rho, a, \lambda), (\rho', a', \lambda')], \lambda') - \vartheta(\rho', a', \lambda')\} \Phi(d\rho, da, d\lambda) \Phi(d\rho', da', d\lambda'), \end{aligned}$$

where $\vartheta(\rho, a, \lambda)$ denotes the current-value co-state variable associated with $\Phi(\rho, a, \lambda)$. In an optimum, the ODEs for the co-state variables are

$$r\vartheta(\rho,a,\lambda)-\dot{\vartheta}(\rho,a,\lambda)=u(\rho,a)+\alpha\int_{-1}^{1}\left(\vartheta(\rho',a,\lambda)-\vartheta(\rho,a,\lambda)\right)dF(\rho')$$

$$+\int_{0}^{1}\int_{-\infty}^{\infty}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}\left\{\vartheta\left(\rho,a+q^{*}\left[(\rho,a,\lambda),(\rho',a',\lambda')\right],\lambda\right)-\vartheta(\rho,a,\lambda)\right.$$

$$\left.+\vartheta\left(\rho',a'-q^{*}\left[(\rho,a,\lambda),(\rho',a',\lambda')\right],\lambda'\right)-\vartheta(\rho',a',\lambda')\right\}\Phi(d\rho',da',d\lambda')$$

s.t.

$$\vartheta_2\left(\rho,a+q^{*}\left[(\rho,a,\lambda),(\rho',a',\lambda')\right],\lambda\right)=\vartheta_2\left(\rho',a'-q^{*}\left[(\rho,a,\lambda),(\rho',a',\lambda')\right],\lambda'\right).$$

It is easy to check that the planner's optimality conditions do not coincide with the equilibrium conditions. More specifically, the comparison with Equation (10) reveals that the planner's optimality conditions and the equilibrium conditions would be identical if there was not $1/2$ in front of the matching rate in the equilibrium condition. This difference is due to a hold-up problem typical of *ex post* bargaining environments. This might seem surprising at first glance because Nash bargaining chooses the efficient trade for each bilateral match. The root cause of inefficiency is the price impact. When negotiating for a trade quantity, investors recognize the fact that their trades will also create a price impact in the future and that the price impact is increasing in the surplus that those future trades generate. Consequently, at the margin, investors tend to avoid taking extreme positions which would probably lead to large price impacts on future trades.

The solution method for the planner's problem is exactly the same as the solution method I used for equilibrium. In the end, the difference between the planner's solution and the equilibrium solution boils down to the use of different effective discount rates. The effective discount rate that the benevolent social planner would assign to investors with λ solves the functional equation

$$\widetilde{r}^*(\lambda) = r + \int_0^1 2\mu\lambda \frac{\lambda'}{\Lambda} \frac{\widetilde{r}^*(\lambda')}{\widetilde{r}^*(\lambda) + \widetilde{r}^*(\lambda')} d\Psi(\lambda'). \tag{32}$$

In the Appendix, I show that $\widetilde{r}^*(\lambda)/\widetilde{r}(\lambda)$ is increasing in λ, implying that fast investors' effective discount rates are distorted to a greater extent. In other words, in equilibrium, fast investors provide immediacy but not as much as the social planner would like them to provide. The quantities chosen by the planner are given by

$$q^*[(\rho, a, \lambda), (\rho', a', \lambda')] = \frac{\frac{\sigma_\eta}{\sigma_D}\left(\frac{\rho' - \overline{\rho}}{\widetilde{r}^*(\lambda') + \alpha} - \frac{\rho - \overline{\rho}}{\widetilde{r}^*(\lambda) + \alpha}\right) + \left(\frac{a' - A}{\widetilde{r}^*(\lambda')} - \frac{a - A}{\widetilde{r}^*(\lambda)}\right)}{\frac{1}{\widetilde{r}^*(\lambda)} + \frac{1}{\widetilde{r}^*(\lambda')}}. \tag{33}$$

Therefore, the allocation implied by the planner's choices solves the following system of Fourier transforms:

$$0 = -(\alpha + 2\mu\lambda)\widehat{\phi}^*_{\rho,\lambda}(z) + \alpha \int_{-1}^1 \widehat{\phi}^*_{\rho',\lambda}(z)\, dF(\rho') \tag{34}$$

$$+ \int_0^1 \int_{-1}^1 2\mu\lambda \frac{\lambda'}{\Lambda} e^{i2\pi \overline{C}^*[(\rho,\lambda),(\rho',\lambda')] \frac{z}{1 + \frac{\widetilde{r}^*(\lambda')}{\widetilde{r}^*(\lambda)}}} \widehat{\phi}^*_{\rho,\lambda}\left(\frac{z}{1 + \frac{\widetilde{r}^*(\lambda')}{\widetilde{r}^*(\lambda)}}\right) \widehat{\phi}^*_{\rho',\lambda'}\left(\frac{z}{1 + \frac{\widetilde{r}^*(\lambda')}{\widetilde{r}^*(\lambda)}}\right) dF(\rho')\, d\Psi(\lambda')$$

for all $\lambda \in \text{supp}(\Psi)$, $\rho \in \text{supp}(F)$ and for all $z \in \mathbb{R}$;

$$\widehat{\phi}^*_{\rho,\lambda}(0) = 1 \tag{35}$$

for all $\lambda \in \text{supp}(\Psi)$ and $\rho \in \text{supp}(F)$; and

$$\int_0^1 \int_{-1}^1 \left(\widehat{\phi}^*_{\rho,\lambda}\right)'(0) dF(\rho)\, d\Psi(\lambda) = -i2\pi A, \tag{36}$$

where

$$\overline{C}^{*}\left[\left(\rho,\lambda\right),\left(\rho',\lambda'\right)\right] \equiv \widetilde{r}^{*}\left(\lambda'\right)\frac{\sigma_{\eta}}{\sigma_{D}}\left(\frac{\rho-\overline{\rho}}{\widetilde{r}^{*}\left(\lambda\right)+\alpha}-\frac{\rho'-\overline{\rho}}{\widetilde{r}^{*}\left(\lambda'\right)+\alpha}\right)-\left[\frac{\widetilde{r}^{*}\left(\lambda'\right)}{\widetilde{r}^{*}\left(\lambda\right)}-1\right]A. \tag{37}$$

4 Assessing the model's implications

4.1 Average holdings, trade sizes and prices

Using the result of Proposition 3, I derive the average asset holdings, trade sizes, and prices of investors of type (ρ,λ). The results are summarized in the following corollary:

Corollary 1 *The average asset holdings, trade sizes and prices of investors of type* (ρ,λ) *are given by:*

$$\mathbb{E}_{\phi}\left[a\mid\rho,\lambda\right]=\frac{\alpha}{\alpha+2\left(\widetilde{r}\left(\lambda\right)-r\right)}A+\frac{2\left(\widetilde{r}\left(\lambda\right)-r\right)}{\alpha+2\left(\widetilde{r}\left(\lambda\right)-r\right)}\left[A-\frac{\sigma_{\eta}}{\sigma_{D}}\frac{\widetilde{r}\left(\lambda\right)}{\widetilde{r}\left(\lambda\right)+\alpha}\left(\rho-\overline{\rho}\right)\right], \tag{38}$$

$$\mathbb{E}_{\phi}\left[q\mid\rho,\lambda\right]=\frac{\alpha}{\alpha+2\left(\widetilde{r}\left(\lambda\right)-r\right)}\left[-\frac{\widetilde{r}\left(\lambda\right)-r}{\mu\lambda}\frac{\sigma_{\eta}}{\sigma_{D}}\frac{\widetilde{r}\left(\lambda\right)}{\widetilde{r}\left(\lambda\right)+\alpha}\left(\rho-\overline{\rho}\right)\right], \tag{39}$$

$$\mathbb{E}_{\phi}\left[P\mid\rho,\lambda\right]=P^{W}-\frac{\alpha}{\alpha+2\left(\widetilde{r}\left(\lambda\right)-r\right)}\left[\left(\rho-\overline{\rho}\right)\frac{r\gamma\sigma_{D}\sigma_{\eta}}{\widetilde{r}\left(\lambda\right)+\alpha}\left(\frac{3}{4}-\frac{\widetilde{r}\left(\lambda\right)-r}{2\mu\lambda}\right)\right]. \tag{40}$$

The implication of equation (38) is intuitive: The average holding is a decreasing function of correlation ρ. As ρ increases, the hedging benefit of the asset decreases and investors hold less of it. The investor with average correlation holds the per capita supply on average. There are two reasons behind the deviation of average OTC holdings from Walrasian holdings which are derived in Section 3.2: Intensive and extensive margin effects. To understand the intensive margin effect, I first define the "desired OTC holding" as the holding which equates the investor's marginal valuation to the average marginal valuation of the market. The desired OTC holding of an investor of type (ρ,λ) is $A-\frac{\sigma_{\eta}}{\sigma_{D}}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\left(\rho-\overline{\rho}\right)$. This shows the distortion of investors' decisions on the intensive margin, i.e., the desired OTC holding is different from the optimal Walrasian holding. More specifically, the coefficient of current correlation in the desired holding is $\frac{\sigma_\eta}{\sigma_D}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}$ instead of $\frac{\sigma_\eta}{\sigma_D}$. Investors put less weight on their current correlation by scaling down the Walrasian weight as previously shown by the partially centralized models of Gârleanu (2009) and Lagos and Rocheteau (2009). This is because investors want to hedge against the risk of being stuck with undesirable positions for long periods upon the arrival of an idiosyncratic shock. They achieve this specific hedging by distorting their decisions on the intensive margin. To understand the extensive margin effect, note that, in equilibrium, we observe investors who have recently become of type (ρ, λ) but have not had the chance to interact with other investors. On average, these investors hold A, due to the i.i.d. and non-persistence of correlation shocks. Equation (38) shows that the average OTC holding is a linear combination of the desired OTC holding and A. Using this interpretation, the fraction $\frac{\alpha}{\alpha+2(\widetilde{r}(\lambda)-r)}$ can be broadly considered to be a measure of the distortion on the extensive margin. When μ is finite, this fraction is bigger than 0, and this creates the second source of the deviation from Walrasian holding. Hence, investors' average asset positions are less extreme than the Walrasian position because of the intensive and extensive margin effects. This analysis also implies that fast investors hold more extreme positions (exhibiting larger deviation from A) than slow investors on average for two reasons. First, since they are able to trade often, their desired asset positions are more extreme. Second, they are exposed to lower distortion on the extensive margin so that their positions are relatively closer to the desired position.

From equation (39), we see that the average trade size is a decreasing function of correlation ρ. The investor with average correlation has 0 net volume on average. Investors with higher correlations are net sellers, and investors with lower correlations are net buyers on average. Average individual trade sizes are also less extreme compared to Walrasian individual trade sizes, since investors trade less aggressively by putting a lower weight on their current correlation.

Equation (40) reveals that the average price is a decreasing function of correlation ρ. The investor with average correlation faces the Walrasian price on average. Investors with higher correlations face lower prices than the Walrasian price, and investors with lower correlations face higher prices than the Walrasian price. Expected sellers trade at lower prices, and expected buyers trade at higher prices because their need to buy or sell is reflected in the transaction price through the bargaining process. In other words, investors with a stronger need to trade, i.e., with high $|\rho|$, trade at less favorable terms. This implication is consistent with empirical evidence in Ashcraft and Duffie (2007) in the federal funds market.

To sum up, the overall pricing implications of my model come from the decisions on the intensive margin: Investors' average asset positions are less extreme as they put less weight on their current valuation and more weight on their future expected valuation for the asset, compared to the frictionless case. In other words, net suppliers of the asset supply less than the Walrasian market, and net demanders of the asset demand less. However, the overall effect on the aggregate demand is zero, and the mean of the equilibrium price distribution is equal to the Walrasian price.[15] Therefore, my model complements the results of the existing purely decentralized markets model by showing that, once portfolio restrictions are eliminated, the pricing impact of search frictions is low. This result is consistent with the findings of illiquid market models such as Gârleanu (2009) and transaction cost models such as Constantinides (1986). These papers show that infrequent trading and high transaction costs have a first-order effect on investors' asset positions, but only a second-order effect on prices, due to the investors' ability to adjust their asset positions. My model demonstrates that a similar intuition carries over to decentralized markets when there are no restrictions

[15] This result is expected to depend on the quadratic specification of $u(\rho, a)$. Indeed, the average price is unaffected by frictions since the marginal utility flow is linear in type and asset position. On the other hand, a more general intuition is highlighted here: The asset demands of different type of investors are affected differently. Hence, the aggregate demand does not have to be affected significantly.

on holdings.

4.2 Dispersion of marginal valuations and asset positions

Using the result of Proposition 3 evaluated at $n = 2$, I obtain a linear system which pins down the conditional variance of asset positions, $var_{\phi}[a|\lambda]$, for all $\lambda \in \{\lambda_1, ..., \lambda_N\}$. I also derive an equation which relates $var_{\phi}[a|\lambda]$ to the conditional variance of marginal valuations, $var_{\phi}[J_2(\rho, a, \lambda)|\lambda]$. This analysis leads to the following corollary:

Corollary 2 *The conditional variance of marginal valuations,* $var_{\phi}[J_2(\rho, a, \lambda)|\lambda]$, *is decreasing in* λ. *The conditional variance of asset holdings,* $var_{\phi}[a|\lambda]$, *is increasing in* λ.

This corollary establishes the lower variability of marginal valuations for fast investors. The dispersion of marginal valuations among the investors with the same λ stems from the difference in the current hedging need or current asset position. In other words, it stems from the effect of the current marginal utility flow on marginal valuations. As fast investors discount their current marginal utility flow using a higher effective discount rate, we observe lower dispersion in their marginal valuation. This is true even though dispersion created by asset positions is bigger for fast investors. Therefore, for investors who are trying to correct their holdings, fast investors become the natural counterparty since their marginal valuations are always close to the average marginal valuation.

Proposition 1 implies that fast investors trade aggressively according to their counterparties' needs. When they meet a buyer, they sell a lot. When they meet a seller, they buy a lot. This is optimal for fast investors: Deviating from the desired position is less of a concern for them as they have higher effective discount rates. As a result of this, fast investors' positions exhibit large volatility. Figure 2 shows it graphically. At time 0, a fast and a slow investor start trading with the same correlation $\rho = -0.19809 < \overline{\rho} = -0.16$, i.e., both of them have higher taste for the asset than the market average. Thus, on average, both of them maintain a position bigger than the supply $A = 8,740$. We see that the average position of the fast investor is more extereme, which is consistent with our discussion in the last section. As time passes, the two investors bump into other investors randomly chosen from the equilibrium distribution. As anticipated, the fast investor's holding exhibits higher volatility.

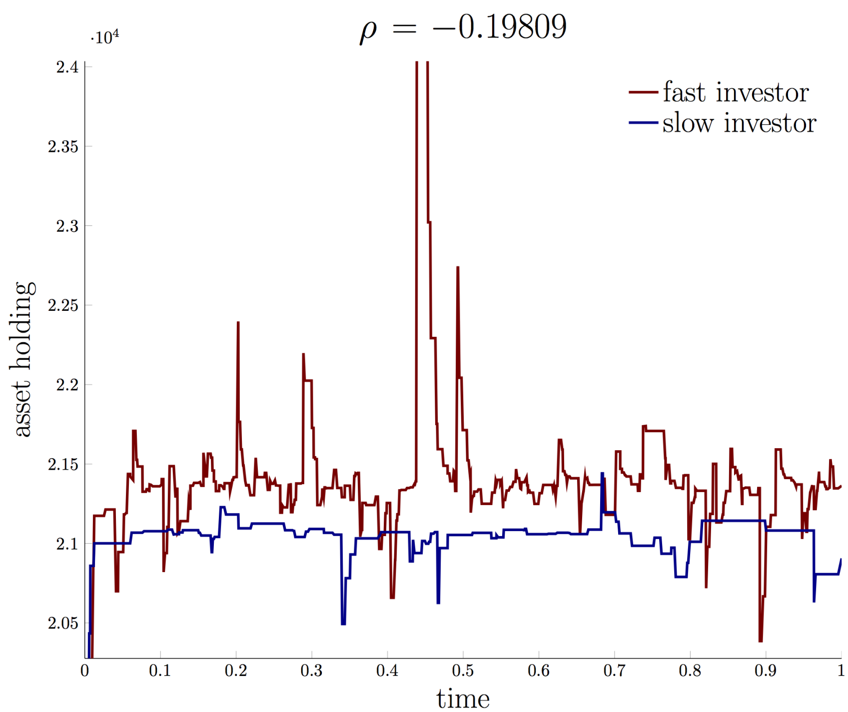

Figure 2. Sample path of asset holdings for two investors with different search intensities

Figure 3 demonstrates the effect of fast investors' volatile inventories on the cross-sectional distribution of asset holdings. The conditional distributions of asset positions for two classes of investors are considered. Both classes have the same correlation type of -1. Thus, these investors are the ones with highest exogenous valuation for the asset. The graph reveals the bimodal structure of both distributions. This stems from the fact that investors with holdings distorted on the extensive margin and investors with average correct holdings create different groups. In the example, investors with holdings distorted on the extensive margin create a group around $A = 8,740$. Slow investors' density is higher around A because the expected period until a trade opportunity after an idiosyncratic shock is higher for them. The second group reflects the fact that the desired holding is different for fast and slow investors. Although both investors like the asset, fast investors hold a higher average position because of the intensive margin effect of the frictions. In addition, we see that fast investors' positions exhibit larger dispersion. This is due to the higher volatility in their positions.

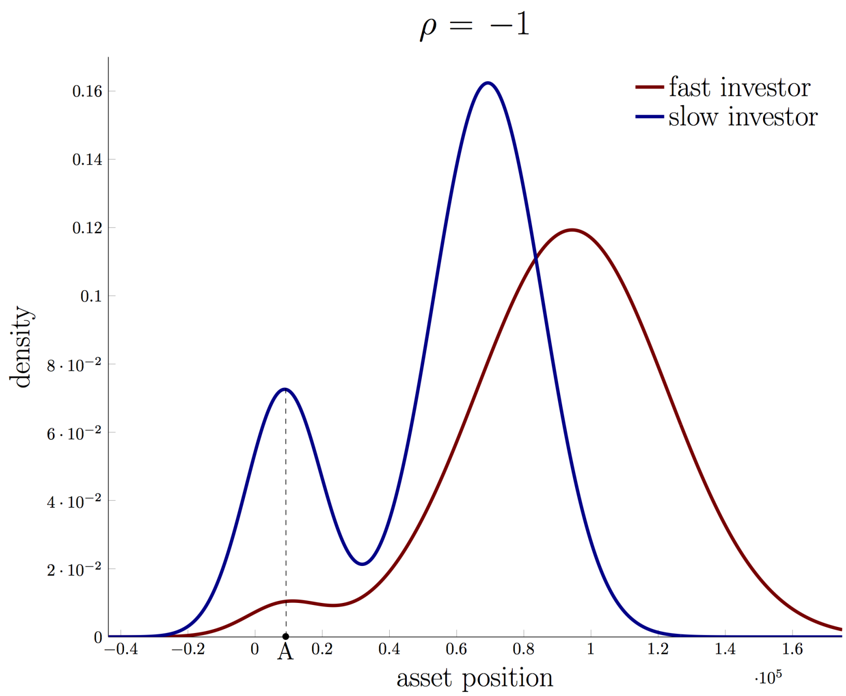

Figure 3. Sample equilibrium conditional distribution of asset holdings for two classes of investors with the same correlation but different search intensities

These results about main intermediation providers holding large and volatile asset positions in equilibrium have important implications for the effects of a section of the Dodd-Frank Act, often referred to as "the Volcker Rule," which disallows proprietary trading by banks and their affiliates. Some forms of proprietary trading are exempted from the Volcker rule, such as those related to market making or hedging. As the equilibrium of my model reveals, even in a stationary world without speculative trading, fast investors hold extreme positions as a result of their optimal hedging behavior, and very volatile positions as a result of market making. Detecting proprieatry trading, which is unrelated to hedging or market making, based on the fluctuations in asset positions would be a very difficult and possibly infeasible task for regulators. Consequently, banks would perceive that they might face a regulatory sanction due to the imperfections of the criteria and metrics that were proposed to detect non-market-making proprietary trading. This would possibly reduce their incentive to provide liquidity. Hence, the elimination of excessive risk-taking by fast investors might come with a reduction in liquidity provision and in the overall quality of risk-sharing as well. In Section 4.5, I will analyze possible scenarios regarding this issue.

4.3 Trading volume

Figure 4 shows the decomposition of individual instantaneous expected trading volume assuming that all investors have the same λ. As the net and gross trading volume, I report $2\mu\lambda\left|\mathbb{E}_{\theta'}[q(\theta,\theta') \mid \theta]\right|$ and $2\mu\lambda\mathbb{E}_{\theta'}[\left|q(\theta,\theta')\right| \mid \theta]$, respectively.[16] Note that, when everyone has the same λ, the sole determinants of trade quantity are the effective types of the trading parties. I label the difference between gross and net trading volume as intermediation volume as it is caused by simultaneous buying and selling instead of fundamental trading. Consistent with the findings of Afonso and Lagos (2015), Atkeson et al. (2015), and Hugonnier et al. (2014), investors with average marginal valuations tend to specialize in intermediation. Their incentive for rebalancing holdings is low. Thus, they engage mostly in simultaneous buying and selling since it leads to profit due to equilibrium price dispersion. However, investors

[16] The characterization of the equilibrium distribution in Proposition 3 allows for the calculation of the usual moments, but not the absolute moments. Due to this technical difficulty, I calculate $\mathbb{E}_{\theta'}\left[\left|q(\theta,\theta')\right| \mid \theta\right]$ numerically only.

with very high or very low marginal valuations engage very little in intermediation as they are mostly concerned with correcting their holding.

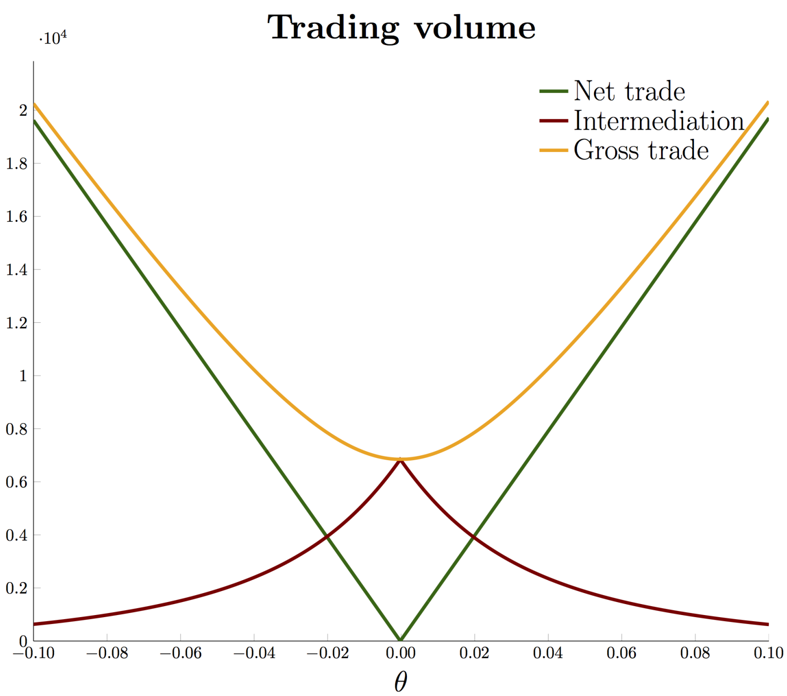

Figure 4. Individual expected instantaneous gross trading volume, net trading volume, and intermediation volume

Since my model features investor heterogeneity together with the unrestricted holdings, it offers a richer explanation of the relation between the investor heterogeneity and the intermediation behavior. Endogenous intermediation models with $\{0, 1\}$ holding such as Hugonnier et al. (2014) and Shen et al. (2015) show that investors with average exogenous valuations specialize as intermediaries. My model offers an alternative explanation with an additional dimension as endogenous asset holding appears to be an important determinant of the marginal valuations. When asset holding is endogenous, having the average marginal valuation means holding the "correct" amount of assets, rather than having the average exogenous valuation. Indeed, as can be seen in Figure 5, any investor with any exogenous valuation can be an intermediary if her holding is "correct". In other words, in my setup with endogenous holdings, intermediaries might be “low valuation-low holding" (red), “average valuation-average holding" (blue), or “high valuation-high holding" (orange) investors.

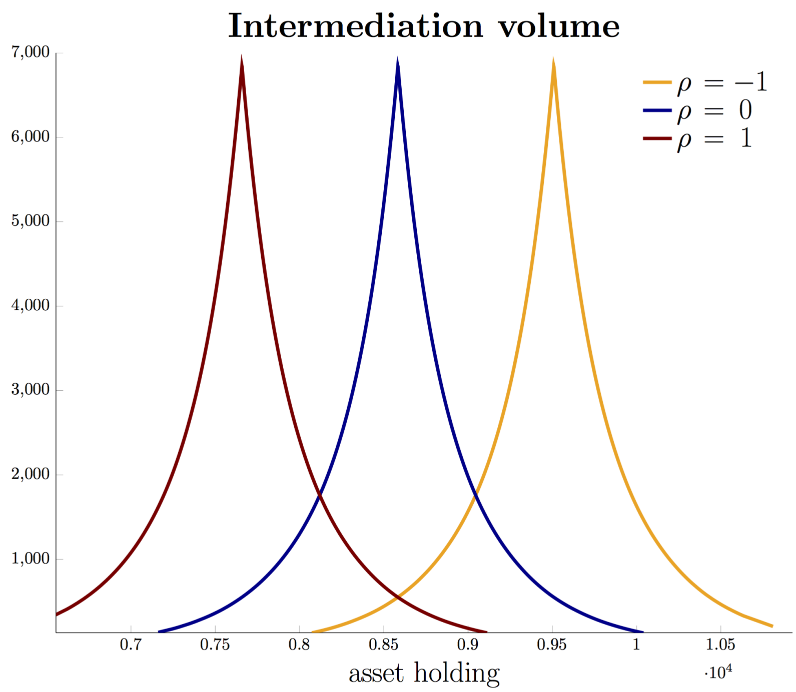

Figure 5. Individual expected instantaneous intermediation volume as a function of asset holding

When I introduce heterogeneity in search intensities, heterogeneity is created in intermediation activity, even controlling for the level of marginal valuation. Fast investors intermediate more due to the effective discount rate channel (see Figure 6). Each bilateral negotiation results in a trade size that is more in line with the slower counterparty’s hedging need, and a trade price that contains a speed premium benefitting the faster counterparty. It is true that fast investors engage in higher simultaneous buying and selling activity due to the higher intensity of matching with counterparties. However, the effective discount rate channel leads to an increase in the intermediation level above that direct effect. Since fast investors trade according to their counterparties' hedging needs, they provide more *intermediation per matching.*

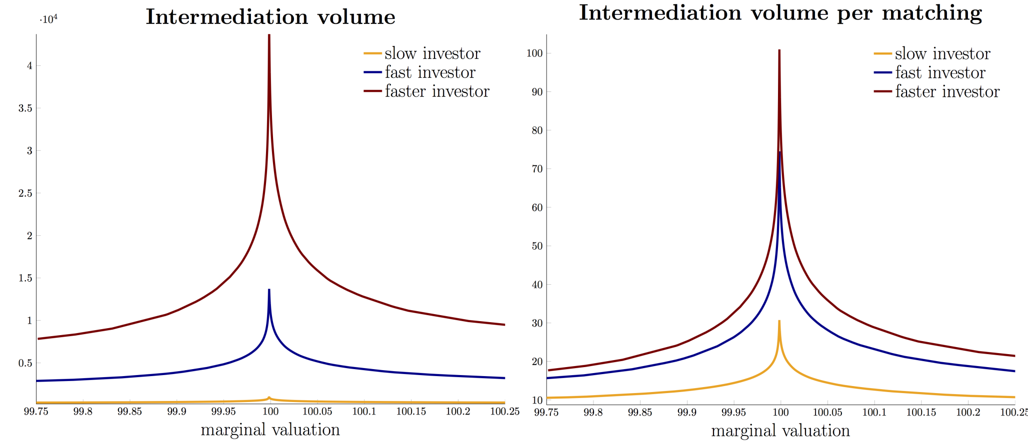

Figure 6. Individual expected instantaneous intermediation volume and intermediation volume *per matching rate* for investors with different search intensities

4.4 An analytical example

In order to derive analytical comparative statics, I focus on a special case of the model with a two-type distribution of search efficiencies. The following lemma provides the closed-form formula for the effective discount rates of the two types of investors.

Lemma 2 *Suppose the support of the distribution, Ψ, is $\{\lambda_s, \lambda_f\}$, where $\lambda_f > \lambda_s$ and ψ_f denotes the fraction of investors with λ_f. Then*

$$\widetilde{r}(\lambda_f) = \begin{cases} \dfrac{-\left(r+\frac{\mu\Lambda}{2}\right)+\left(1-\psi_f\right)\left(r+\frac{\mu\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)+\left(1-\psi_f\right)\sqrt{\left(r+\frac{\mu\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)^2+\frac{\mu\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)}}{1-2\psi_f} & \text{if } \psi_f \neq \frac{1}{2} \\[2ex] \dfrac{\lim\limits_{\psi_f\to\frac{1}{2}} \frac{\partial}{\partial\psi_f}\left\{\left(r+\frac{\mu\Lambda}{2}\right)-\left(1-\psi_f\right)\left(r+\frac{\mu\mathbb{E}\left[\lambda^2\right]}{4\Lambda}+\sqrt{\left(r+\frac{\mu\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)^2+\frac{\mu\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)}\right)\right\}}{2} & \text{if } \psi_f = \frac{1}{2} \end{cases}$$

and

$$\widetilde{r}(\lambda_s) = \begin{cases} \dfrac{r+\frac{\mu\Lambda}{2}-\psi_f\left(r+\frac{\mu\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)-\psi_f\sqrt{\left(r+\frac{\mu\mathbb{E}[\lambda^2]}{4\Lambda}\right)^2+\frac{\mu\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)}}{1-2\psi_f} & \text{if } \psi_f \neq \frac{1}{2} \\ \dfrac{\lim\limits_{\psi_f \to \frac{1}{2}} \frac{\partial}{\partial\psi_f}\left\{-\left(r+\frac{\mu\Lambda}{2}\right)+\psi_f\left(r+\frac{\mu\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)+\psi_f\sqrt{\left(r+\frac{\mu\mathbb{E}[\lambda^2]}{4\Lambda}\right)^2+\frac{\mu\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)}\right\}}{2} & \text{if } \psi_f = \frac{1}{2}. \end{cases}$$

Plugging the effective discount rates given by Lemma 2 into the formulas in Corollary 1, I obtain average equilibrium objects in closed form. Then, I plot some comparative statics graphs.

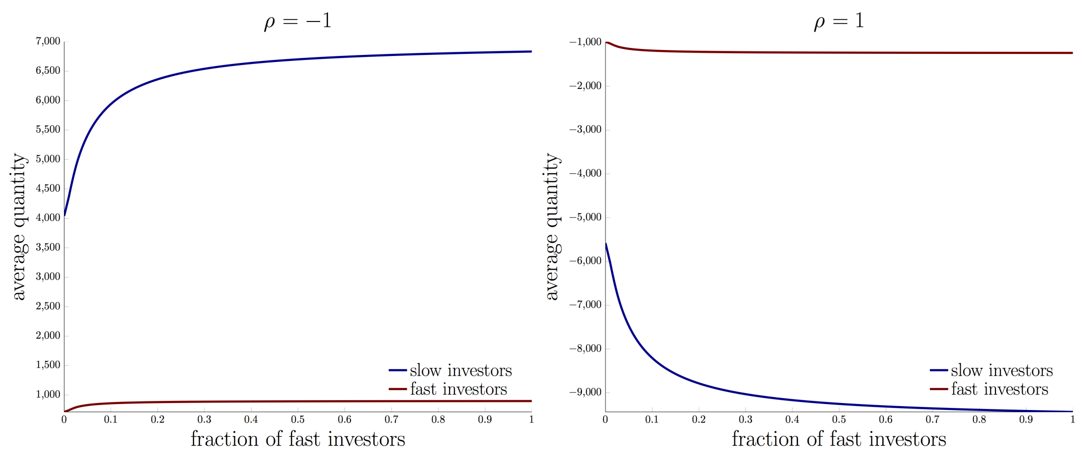

Figure 7. Average net trade quantities as a function of the fraction of fast investors

When we analyze the average net trade quantity (39), we see that there are competing forces. On the one hand, the fast investors with holdings distorted on the extensive margin have high net trade quantities because of liquidity provision incentives and more aggressive trading. The liquidity provision incentive stems from the difference in search intensities.[17] When two buyers with the same correlation type but different search intensities meet, the

[17] For an arbitrary distribution of search intensities, the behavior of $\frac{\widetilde{r}(\lambda)-r}{\mu\lambda}$ is not clear. However, for the two-type distribution, this object is higher for the fast investor.

fast investor will provide liquidity to the slow investor by buying or selling the asset. The more aggressive trading is due to the fact that the high search intensities make the investor less afraid of being stuck with an undesirable position in the future. On the other hand, high search intensity reduces the average net trade quantity by reducing the distortion on the extensive margin. In the two-type case, the latter effect dominates and the average net trade quantity of fast investors is lower. Figure 7 shows the comparative statics with respect to the fraction of fast investors.

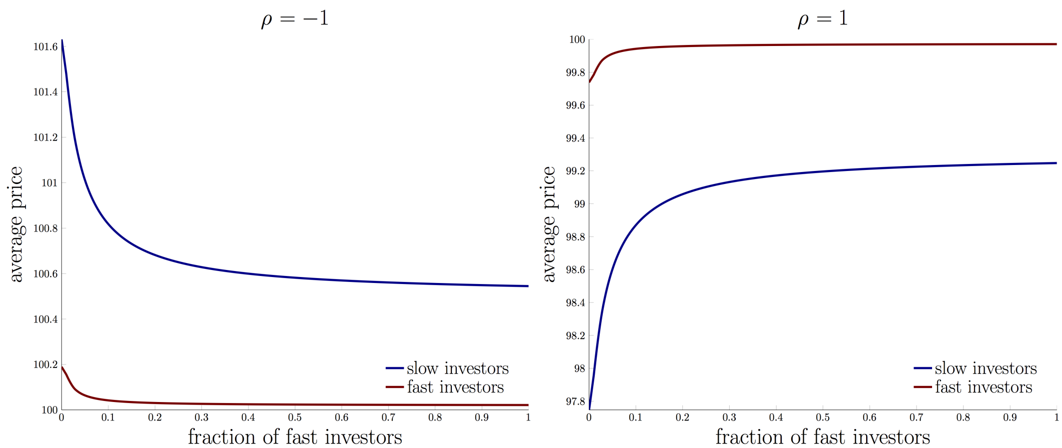

Figure 8. Average prices as a function of the fraction of fast investors

When we look at the average price (40), we see that it is a decreasing function of correlation ρ. The group of investors with the correct holding on average faces the Walrasian price on average. Investors with misallocated holdings face lower prices than the Walrasian price if they have high correlation types, and face higher prices if they have low correlation types. In other words, investors with a stronger need to trade, i.e., with high $|\rho|$, trade at less favorable terms. We see that the investor's λ affects the deviation term from the Walrasian price through three channels. First, since the measure of distortion on the extensive margin is lower for high λ investors, a high fraction of them trade at the Walrasian price on average. Second, since their marginal valuation does not depend much on their current marginal utility flow, their need to trade is reflected by the price to a lesser extent. Finally, there is the effect of the speed premium. Because of these three factors, high λ investors' average trade price is closer to the Walrasian price, while the average trade price of low λ investors deviates a lot. Figure 8 shows the comparative statics with respect to the fraction of fast investors in the two-type case. In the example, the Walrasian price is 100. As the fraction of fast investors increases, both buyers' and sellers' average price becomes closer to the Walrasian price, reflecting the increase in liquidity. As overall liquidity increases, the average speed premium, reflected by the difference between the slow and fast investors' average price, decreases. This is intuitive because when there are more fast traders in the market, slow traders' outside option is closer to the average marginal valuation of the market, lowering the trade surplus, and, in turn the speed premium. In other words, fast investors are able to charge higher speed premia when they only constitute a concentrated, small part of the market.

4.5 A numerical example

In this section, I present a numerical example of my model to capture the heterogeneity among intermediaries observed in the secondary market for municipal bonds, a typical example of OTC markets. The purpose of this exercise is to illustrate that, once I calibrate the model to match certain features of the municipal bond market, the model generates quantitatively meaningful results in terms of trade sizes and the relationship between centrality and intermediation markups. Table 1 shows the parameter values chosen for the baseline calibration.

Since the preference structure of my model is same as that of Duffie et al. (2007), I follow them in setting the discount rate to 5% and the risk aversion parameter to 0.01. I normalize the asset supply to $A = 8,740$ so that the average equilibrium price is $\mathbb{E}_\phi[P] = 100$.[18] The expected asset payoff, $m_D = 5.12$, and the volatility of asset payoff, $\sigma_D = 1.63$, are chosen to match the average yearly excess return of $m_D/\mathbb{E}_\phi[P] - r = 0.12\%$ and the standard deviation of yearly returns of $\sigma_D/\mathbb{E}_\phi[P] = 1.63\%$ (Green, Hollifield & Schürhoff, 2007).

Table 1: Parameter values

Parameter		Value
Discount rate	r	0.05
Risk aversion	γ	0.01
Expected asset payoff	m_D	5.12
Volatility of asset payoff	σ_D	1.63
Volatility of background risk	σ_η	88000
Asset supply	A	8740
Aggregate search efficiency	μ	500
Number of search efficiency types	I	50
Search efficiencies	λ_i	$\beta^{-1}_{(0.0625,0.625)}(0.05 + 0.90[i-1]/I)$
Intensity of idiosyncratic shocks	α	0.125
Number of correlation types	J	10
Correlation types	ρ_j	$-1 + 2\beta^{-1}_{(2000,3000)}([j-1]/J)$

$\beta^{-1}_{(\alpha,\beta)}(x)$ refers to the inverse cumulative function of a beta distribution with an alpha parameter of α and a beta parameter of β.

The aggregate search efficiency, $\mu = 500$, implies a transaction frequency $\frac{\mu\Lambda}{250} = 0.117$ per day, which is in the range of $0.04 - 0.12$ reported by Green et al. (2007). Since my model features a continuum of investors, the expected number of links (degree centrality) and the expected number of trades coincide. I target the size-weighted degree centralities to capture the fact that the network of trade is not random although matching is random. Hence, any given distribution Ψ of search intensities implies a certain distribution of degree centralities. I choose Ψ to match roughly the empirical degree centrality distribution of

[18] When I scale up (down) m_D and σ_D, and scale down (up) A by the same constant, all equilibrium objects I calculate for my numerical exercise stay the same. That is, if $\{q, P, \Phi(\rho, a, \lambda)\}$ is an equilibrium when the asset supply is A, the expected asset payoff is m_D and the asset payoff volatility is σ_D, then, for any $k > 0$, $\left\{\frac{q}{k}, kP, \frac{1}{k}\Phi(\rho, ka, \lambda)\right\}$ is an equilibrium when the asset supply is $\frac{A}{k}$, the expected asset payoff is km_D, and the asset payoff volatility is $k\sigma_D$.

the municipal bonds market that Li and Schürhoff (2012) report. Figure 9a shows the distribution of degree centralities implied by the calibrated distribution of search intensities, and the empirical distribution of degree centralities.

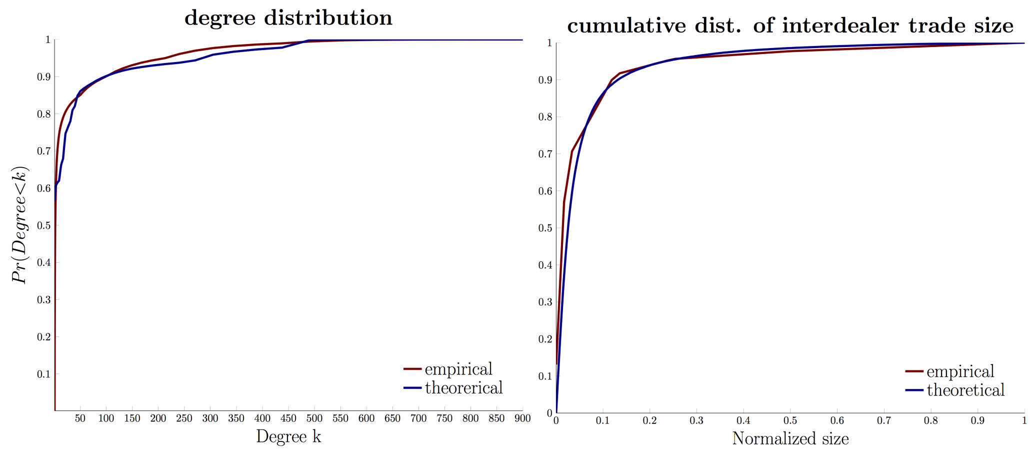

Figure 9. (a) CDF of degree centralities (b) CDF of normalized interdealer trade sizes

In the calibration of the idiosyncratic shocks in my numerical example, I target the asset turnover and the relative distribution of interdealer trade sizes observed in the municipal bond market. In the model, to identify "dealers," I calculate, for any investor, the intermediation volume as a fraction of her gross volume. I label as dealers the investors with the highest intermediation share whose trade among themselves accounts for 21.5% of all trades.[19] The chosen intensity of idiosyncratic shocks, $\alpha = 0.125$, leads to a turnover of 59.1% per year. The counterpart in the data is around 56% per year (Green et al., 2007). The calibration target for the distribution of of correlation types, F, is the empirical distribution of relative interdealer trade sizes. Figure 9b shows the trade size distribution generated by the calibrated distribution of correlation types, and the trade size distribution reported

[19] Li and Schürhoff (2012) document that around 21.5% of trades in the municipal bond market take place between dealers.

by the Fact Book of the Municipal Securities Rulemaking Board (2008).[20]

Finally, for the calibration of the volatility of background risk, σ_η, I target the average intermediation markup during the 1998-2012 period, which Li and Schürhoff (2012) measured. Calculation of the theoretical counterpart of the markups and related moments is described in Appendix D. In my numerical example, the average intermediation markup turns out to be 2.41%, which is in the ballpark of the empirical range of $1.85\% - 2.08\%$.

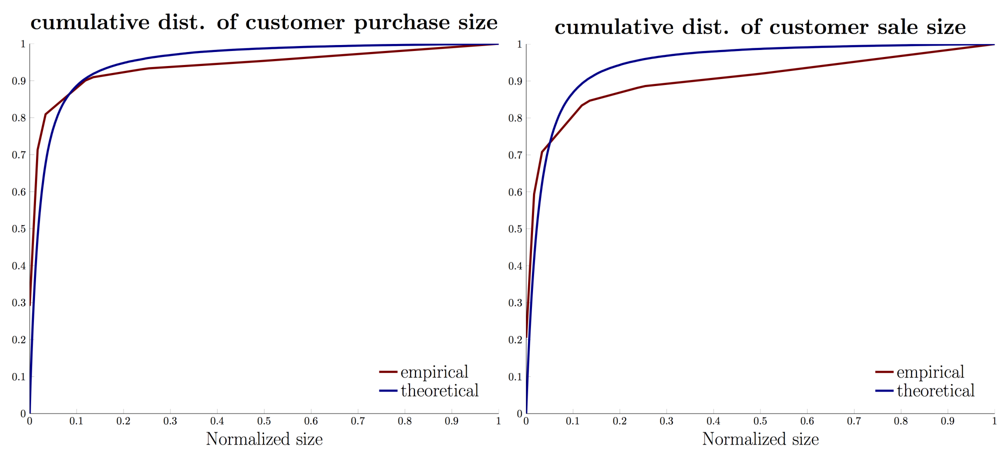

Figure 10. (a) CDF of normalized customer purchase sizes (b) CDF of normalized customer sale sizes

To test the quantitative success of the model, I look at the OLS beta of markup on degree centrality and the distribution of relative trade sizes for trades between dealers and customers. The regressions of Li and Schürhoff (2012) imply that the OLS beta of markup on degree centrality would be between $0.01 - 0.012$.[21] The OLS beta of markup on degree

[20] The trade sizes are unbounded in my model. In the normalization, I choose the trade size that corresponds to the 99^{th} percentile as the maximum trade size.

[21] Li and Schürhoff (2012) run regressions of the markup on an aggregated network measure (the average of various network measures, such as degree centrality, size-weighted degree centrality, cliquishness etc.) and report a beta of about 0.5. Their aggregated network measure and degree centrality are almost perfectly correlated. Their aggregated network measure ranges between -1.721 and 18.868, with a standard deviation of 2.248, while the size-weighted degree centrality ranges between 0 and 4164, with a standard deviation of 93.79. Assuming perfect correlation between their aggregated network measure and the degree centrality,

centrality in my numerical example is 0.0067, which is smaller than but comparable to the counterpart measured in the data. Figure 10 compares the trade sizes generated by the model and observed in the municipal bond market based on the Fact Book of the Municipal Securities Rulemaking Board (2008).

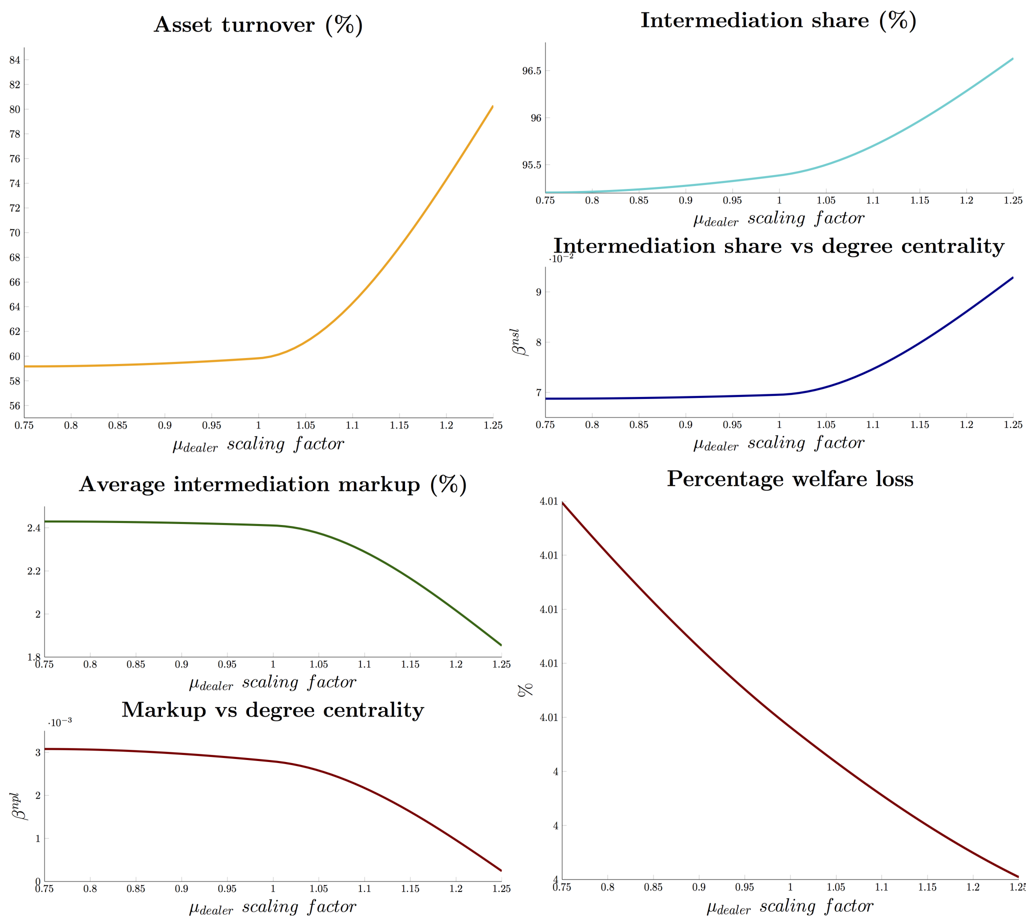

Figure 11. Impact of the search intensity of the main intermediaries

on aggregate outcomes

---

a normalization based on range and standard deviation implies that the OLS beta of markup on degree centrality should be between 0.01 and 0.012.

I conclude this section with comparative statics concerning the dealer's search intensity. In this exercise, I multiply the dealers' search intensity by a scaling factor while I keep the customers' search intensity fixed. I look at how various aggregate moments, welfare, and the distribution of welfare among different investors change. Then I explain how this exercise informs the debate on the Volcker rule. Figure 11 shows the related graphs of aggregate outcomes.

When we look at Figure 11, we first notice that the asset turnover increses with dealers' search intensity. This is a result of the interaction of three distinct effects. First, the total matching rate of a fraction of market participants increases, and this leads to a direct increase in the trading volume. Second, the effective discount rates increase, leading to less cautious trading behavior on the intensive margin, and, hence, creating a positive impact on trading volume. On the other hand, the higher matching rate decreases the distortion on the extensive margin and reduces the need to trade in equilibrium. The first two effects dominate, and the trading volume and, in turn, the asset turnover increase. When we look at the intermediation volume as a fraction total volume, we see that the model predicts an increase. Since the trading volume increases and the distortion on the extensive margin decreases at the same, most of the trades take place for intermediation purposes. To examine the effect of the increase in the dealers' search intensity on the intermediation share of different investors, I run the following regression:

$$\frac{intermediation\ share_i}{average\ intermediation\ share} = \beta^{nsl}\left(degree\ centrality_i\right) + \epsilon_i^{nsl}.$$

This measures the relative importance of fast investors as suppliers of liquidity. β^{nsl} is positive, implying that fast investors have higher intermediation shares. An increase in the main intermediaries' search intensities widens the difference in the effective discount rates between slow and fast investors. Then we observe the further concentration of intermediation activity in the hands of those main intermediaries. Thus, the model predicts an increase in β^{nsl}. On the other hand, the increase in the dealers' search intensity leads to a decline in the intermediation markups due to lower price dispersion caused by higher efficiency of asset allocation. Again, to understand the effect of this change across investors, I run the following regression:

$$\frac{markup_i}{average\ markup} = \beta^{npl}\left(degree\ centrality_i\right) + \epsilon_i^{npl}.$$

This measures the centrality premium or discount in the intermediation markups. The positive β^{npl} demonstrates that fast investors earn higher markups. As dealers' search intensity increases, the competition among dealers becomes stronger and reduces β^{npl}. When we look at the percentage welfare loss graph, we see that the welfare loss caused by OTC market frictions is around 4% of the constrained efficient welfare in consumption equivalent terms. The reduction in the welfare loss stems from the improvement in the efficiency of asset allocation. However, the impact on the distribution of welfare among different investors is not trivial. Figure 12 shows that decomposition.

The fundamental sources of welfare in this environment are the allocation of dividend risk and the hedging benefit. Therefore, by assuming that the alloaction of assets at date 0 is the steady-state equilibrium allocation, the welfare that an investor with λ creates by participating in this market is defined as

$$\mathbb{W}\left(\lambda\right) = \int_0^\infty e^{-rt}\left\{\int_{-\infty}^{\infty}\int_{-1}^{1} u\left(\rho,a\right)\Phi_{\lambda,t}\left(d\rho,da\right)\right\}dt = \frac{1}{r}\int_{-\infty}^{\infty}\int_{-1}^{1} u\left(\rho,a\right)\Phi_{\lambda}\left(d\rho,da\right).$$

Derivations in Appendix E imply that

$$
\begin{aligned}
\mathbb{W}(\lambda) = \frac{m_D}{r}A - \frac{\gamma\sigma_D^2}{2}A^2 - \gamma\sigma_D\sigma_\eta\overline{\rho}A \\
&+ \gamma\sigma_\eta^2 var[\rho]\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\left(\frac{2(\widetilde{r}(\lambda)-r)}{\alpha+2(\widetilde{r}(\lambda)-r)} - \frac{1}{2}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\right) \\
&- \frac{\gamma\sigma_D^2}{2}\frac{r}{\widetilde{r}(\lambda)}\left(var_\phi[\theta|\lambda] - \frac{2\alpha}{\alpha+2(\widetilde{r}(\lambda)-r)}\left(\frac{\sigma_\eta}{\sigma_D}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\right)^2 var[\rho]\right).
\end{aligned}
$$

From this equation, we see that a change in the dealers' search intensity affects the welfare through the last two terms (terms on the last two lines). The first term captures the hedging benefit net of the cost of fundamental risk-bearing. The second term captures the discounted value of the additional cost of (or reduction in) risk-bearing due to heterogeneity in search intensities. This second term stems from the expected deviation of an investor from her desired asset position throughout her lifetime due to trades with investors with different effective discount rates. In a market where every investor has the same λ, this term would be 0. Numerical analysis (see Figure 12) shows that there exists a $\overline{\lambda}$ such that the second term is negative (positive) for $\lambda < \overline{\lambda}$ ($\lambda > \overline{\lambda}$). In other words, the second term creates a cost for fast investors while it creates a benefit for slow investors.

As the dealers' search intensity increases, the effective discount rates increase. The increase in the effective discount rates reduces the distortion on both the extensive and intensive margins. When we look at the first term, we see that the improvement on the extensive margin increases welfare due to an increased hedging benefit, while the improvement on the intensive margin decreases welfare due to increased risk-taking. A simple first derivative analysis implies that, for the first term, the extensive margin dominates and we observe an increase in the first term. The impact on the second term is more complicated since different forces dominate for different investors. A key observation is that the instantaneous matching rate of slow investors stays the same while the dealers' matching rate increases. Therefore, the reduction in the distortion on the extensive margin for slow investors is not strong enough. On the other hand, conditional on a matching, the probability of trading with a fast investor increases. The widening of the difference between the effective discount rates of fast and slow investors increases the extent to which the trade quantities reflect the slow investors' hedging need. Therefore, for slow investors, the reduction in the distortion on the intensive margin is substantial, and this is the dominant impact. In other words, although slow investors get close to their desired position faster, their more aggressive trading behavior leaves them with more undesirable positions after an idiosyncratic shock. Consequently, the increase in dealers' search intensity leads to an increase in the risk-taking due to heterogeneity, and we observe an increase in the second term for slow investors. For fast investors, the story is the exact opposite. Since their instantaneous matching rate increases, there is a strong reduction in the distortion on the extensive margin for fast investors. Therefore, this dominates the effect of the reduction in the distortion on the intensive margin, and we observe a decrease in the second term for fast investors. Therefore, slow investors' benefit from the second term decreases, while fast investors' cost from the second term decreases, leading to a flattening of the second term along the λ dimension. Combined with the first term, this results in a welfare loss for sufficiently slow investors and a welfare gain for others.[22]

Summing over the welfare created by all investors, the social welfare is

[22] Note that the welfare creation and the actual welfare received by an individual are not the same thing, due to the transfer of the numéraire among investors when they trade. Accordingly, an additional reason why slow investors' welfare decreases is that fast investors are able to extract higher surplus from slow investors through the speed premium. This can be seen in Figure 12. Although we observe a reduction in welfare creation for only a small fraction of slow investors, a higher fraction of investors have lower welfare as the dealers' search intensity increases.

$$
\begin{aligned}
\mathbb{W} = {} & \frac{m_D}{r} A - \frac{\gamma \sigma_D^2}{2} A^2 - \gamma \sigma_D \sigma_\eta \overline{\rho} A \\
& + \gamma \sigma_\eta^2 var\left[\rho\right] \int_0^1 \frac{\widetilde{r}\left(\lambda\right)}{\widetilde{r}\left(\lambda\right)+\alpha} \left[\frac{2\left(\widetilde{r}\left(\lambda\right)-r\right)}{\alpha + 2\left(\widetilde{r}\left(\lambda\right)-r\right)} - \frac{1}{2} \frac{\widetilde{r}\left(\lambda\right)}{\widetilde{r}\left(\lambda\right)+\alpha}\right] d\Psi\left(\lambda\right) \\
& - \frac{\gamma \sigma_D^2}{2} \int_0^1 \frac{r}{\widetilde{r}\left(\lambda\right)} \left(var\left[\theta | \lambda\right] - \frac{2\alpha}{\alpha + 2\left(\widetilde{r}\left(\lambda\right)-r\right)} \left(\frac{\sigma_\eta}{\sigma_D} \frac{\widetilde{r}\left(\lambda\right)}{\widetilde{r}\left(\lambda\right)+\alpha}\right)^2 var\left[\rho\right]\right) d\Psi\left(\lambda\right).
\end{aligned}
$$

The impact of the increase in the dealers' search intensity on the social welfare can also be understood from the tension between the last two terms. The first term increases in response to an increase in the dealers' search intensity, as it increases for all λs. My numerical analysis shows that the second term also increases because the flatenning of the second term is higher for slow investors. If the first term is dominant, the welfare increases. The relative sensitivity of the first term to effective discount rates is higher when α is high and μ is low. Therefore, in markets with a positive relationship between centrality and markups (i.e., in markets with high α and low μ), the first term turns out to be dominant, while the second term is dominant in markets with a negative relationship between centrality and markups.

These results have implications for the Volcker Rule. Duffie (2012b) says that "the market making is inherently a form of proprietary trading. A market maker acquires a position from its client at one price and then lays off the position over time at an uncertain average price" (p. 3). He continues by arguing that banning proprietary trading would effectively make offering market making unattractively risky for banks. Following his arguments, I assume that, under the Volcker Rule, the key intermediaries' incentive to act as intermediaries would be reduced. Accordingly, in my model, I capture this in a stylized way by decreasing the dealers' measure of specialists. As discussed above, my model predicts different welfare impacts for different markets. While it would be beneficial for markets with a negative relation between centrality and markups (e.g. the market for asset-backed securities), it would be harmful for markets with a positive relation between centrality and markups (e.g. the municipal bond market).

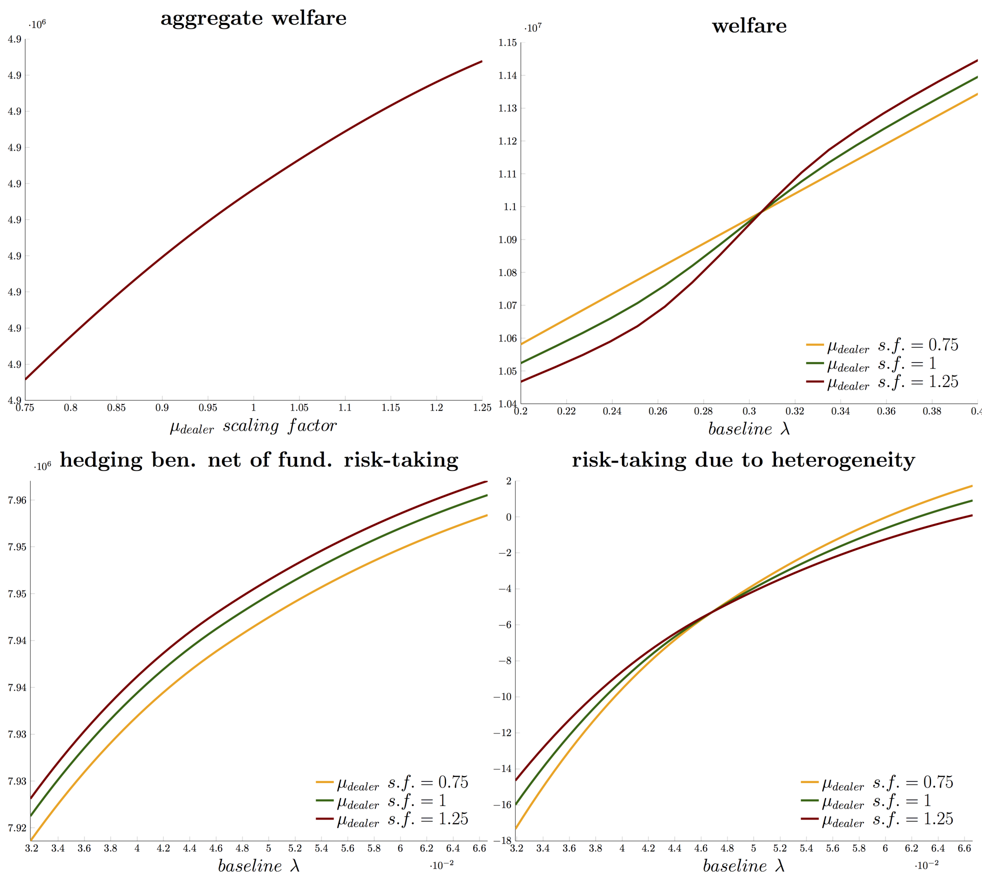

Figure 12. Impact of the search intensity of the main intermediaries on the distribution of welfare

5 Conclusion

OTC markets played a significant role in the recent 2007-2008 financial crisis, as derivative securities, collateralized debt obligations, repurchase agreements, and many other assets are traded OTC. Accordingly, understanding the functioning of these markets, detecting potential inefficiencies, and proposing regulatory action have become a focus of attention for economists and policy makers. This paper contributes to a fast-growing body of literature on OTC markets by presenting a search-and-bargaining model *à la* Duffie et al. (2005). I complement this literature by considering investors who can differ in their search intensities, time-varying hedging needs, and asset holdings. By means of its rich heterogeneity, my model accounts for many observed trading patterns in OTC markets. Investors with higher search intensities (i.e., fast investors) arise endogenously as the main intermediation providers. Then, as observed in the data, they hold large and volatile inventories. Depending on the level of frictions, they can earn higher or lower markups than slow investors. Both are observed in real-life OTC markets. The model's insight into the relation between frictions and the sign of the relation between centrality and markups has further implications in terms of welfare. The consistency between the model outcomes and the positive facts gives us confidence with regard to trusting these welfare implications. Using parametric examples of my model, I show that the regulations that aim to limit the role of central intermediaries, such as the Volcker rule, would have adverse welfare impact on markets with high levels of frictions, while they would be beneficial in markets with low levels of frictions.

This paper leads to several avenues for future research. First, I consider a stationary equilibrium in this paper. Intermediation becomes especially important at times of financial distress. To this end, I plan to study the transitional dynamics of endogenous intermediation, following an aggregate liquidity shock. The dynamics of the price and supply of liquidity along the recovery path could inform the debate on optimal policy during crises. Second, this paper presents a single-asset model. I plan to analyze how endogenous intermediation patterns change in a setup with multiple assets. This analysis could lead to interesting dynamics of liquidity across markets, as maintaining high inventory in one market would limit an intermediary's ability to provide liquidity in other markets. Finally, this paper is totally agnostic about why we observe an *ex ante* heterogeneity in search intensity. Given that this search heterogeneity is an important source of intermediation, studying a model with endogenous search intensities would be a worthwhile way to explore whether the size of the intermediary sector is socially efficient.

Appendix A. Microfoundations for the mean-variance utility flow

Assume that there are two assets. One asset is riskless and pays interest at an exogenously given rate r. This asset is traded in a continuous frictionless market. The other asset is risky, traded over the counter, and is in supply denoted by A. This asset pays a cumulative dividend:

$$dD_t = m_D dt + \sigma_D dB_t, \tag{A.1}$$

where B_t is a standard Brownian motion.

I borrow the specification of preferences and trading motives from Duffie et al. (2007) and Gârleanu (2009). Investors are subjective expected utility maximizers with CARA felicity functions. Investors' coefficient of absolute risk aversion and time preference rate are denoted by γ and r respectively.

Investor i has cumulative income process η^i:

$$d\eta_t^i = m_\eta dt + \sigma_\eta dB_t^i, \tag{A.2}$$

where

$$dB_t^i = \rho_t^i dB_t + \sqrt{1 - (\rho_t^i)^2} dZ_t^i. \tag{A.3}$$

The standard Brownian motion Z_t^i is independent of B_t, and ρ_t^i captures the instantaneous correlation between the payoff of the risky asset and the income of investor i. This correlation is time-varying and heterogeneous across investors. Thus, this heterogeneity creates the gains from trade. In the context of different markets, this heterogeneity can be interpreted in different ways such as hedging demands or liquidity needs.

I assume that the correlation between an investor's income and the payoff of risky asset is itself stochastic. Stochastic processes that govern idiosyncratic shocks and trade are as described in Section 2.

Let $V(W, \rho, a, \lambda)$ be the maximum attainable continuation utility of investor of type (ρ, a, λ) with current wealth W. It satisfies

$$V(W, \rho, a, \lambda) = \sup_c \mathbb{E}_t \left[-\int_t^{\infty} e^{-r(s-t)} e^{-\gamma c_s} ds \mid W_t = W,\ \rho_t = \rho,\ a_t = a \right], \tag{A.4}$$

s.t. $$dW_t = (rW_t - c_t)dt + a_t dD_t + d\eta_t - P\left[(\rho_t, a_t, \lambda), (\rho'_t, a'_t, \lambda'_t)\right] da_t$$

$$da_t = \begin{cases} q\left[(\rho_t, a_t, \lambda), (\rho'_t, a'_t, \lambda'_t)\right] & \text{if there is contact with investor } (\rho'_t, a'_t, \lambda'_t) \\ 0 & \text{if no contact,} \end{cases} \tag{A.5}$$

where $\{q\left[(\rho, a, \lambda), (\rho', a', \lambda')\right], P\left[(\rho, a, \lambda), (\rho', a', \lambda')\right]\} =$

$$\arg\max_{q,P} \left[V(W - qP, \rho, a + q, \lambda) - V(W, \rho, a, \lambda)\right]^{\frac{1}{2}} \left[V(W' + qP, \rho', a' - q, \lambda') - V(W', \rho', a', \lambda')\right]^{\frac{1}{2}}, \tag{A.7}$$

$$\text{s.t. } V(W - qP, \rho, a + q, \lambda) \geq V(W, \rho, a, \lambda),$$
$$V(W' + qP, \rho', a' - q, \lambda') \geq V(W', \rho', a', \lambda').$$

Since investors have CARA preferences, terms of trade are independent of wealth levels as I will show later. To eliminate Ponzi-like schemes, I impose the transversality condition

$$\lim_{T \to \infty} e^{-r(T-t)} \mathbb{E}_t \left[e^{-r\gamma W_T} \right] = 0. \tag{A.8}$$

To derive the optimal rules, the technique of stochastic dynamic programming is used following Merton (1971). Assuming sufficient differentiability and applying Ito's lemma for jump-diffusion processes, the investor's value function $V(W, \rho, a, \lambda)$ satisfies the Hamilton-Jacobi-Bellman (HJB) equation

$$
\begin{aligned}
0 = \sup_c \{ & -e^{-\gamma c} + V_W(W,\rho,a,\lambda)[rW - c + am_D + m_\eta] \\
& + \frac{1}{2} V_{WW}(W,\rho,a,\lambda)[\sigma_\eta^2 + 2\rho a \sigma_D \sigma_\eta + a^2 \sigma_D^2] \\
& - rV(W,\rho,a,\lambda) + \alpha \int_{-1}^{1} [V(W,\rho',a,\lambda) - V(W,\rho,a,\lambda)] dF(\rho') \\
& + \int_{-\infty}^{\infty} \int_{-1}^{1} \{ V\left(W - q\left[(\rho,a,\lambda),(\rho',a',\lambda')\right] P\left[(\rho,a,\lambda),(\rho',a',\lambda')\right], \rho, a + q\left[(\rho,a,\lambda),(\rho',a',\lambda')\right], \lambda\right) \\
& - V(W,\rho,a,\lambda) \} \, 2\mu\lambda \frac{\lambda'}{\Lambda} \Phi(d\rho', da', d\lambda') \}. \qquad \text{(A.9)}
\end{aligned}
$$

Following Duffie et al. (2007), I guess that $V(W, \rho, a, \lambda)$ takes the form

$$V(W, \rho, a) = -e^{-r\gamma(W + J(\rho,a,\lambda) + \overline{J})} \qquad \text{(A.10)}$$

for some function $J(\rho, a)$, where

$$\overline{J} = \frac{1}{r}\left(m_\eta + \frac{\log r}{\gamma} - \frac{1}{2} r\gamma\sigma_\eta^2 \right) \qquad \text{(A.11)}$$

is a constant. Replacing into (A.9), I find that the optimal consumption is

$$c = -\frac{\log r}{\gamma} + r(W + J(\rho, a, \lambda) + \overline{J}).$$

After plugging c back into (A.9) and dividing by $r\gamma V(W, \rho, a, \lambda)$, I find that (A.9) is satisfied iff

$$rJ(\rho,a,\lambda) = u(\rho,a) + \alpha\int_{-1}^{1}\frac{1-e^{-r\gamma[J(\rho',a,\lambda)-J(\rho,a,\lambda)]}}{r\gamma}dF(\rho')$$
$$+\int_{-\infty}^{\infty}\int_{-1}^{1}\frac{1-e^{-r\gamma\{J(\rho,a+q[(\rho,a,\lambda),(\rho',a',\lambda')],\lambda)-J(\rho,a,\lambda)-q[(\rho,a,\lambda),(\rho',a',\lambda')]P[(\rho,a,\lambda),(\rho',a',\lambda')]\}}}{r\gamma}$$
$$2\mu\lambda\frac{\lambda'}{\Lambda}\Phi(d\rho',da',d\lambda'). \quad \text{(A.12)}$$

Terms of individual trades, $q\left[(\rho,a,\lambda),(\rho',a',\lambda')\right]$ and $P\left[(\rho,a,\lambda),(\rho',a',\lambda')\right]$, are determined by a Nash bargaining game with the solution given by the optimization problem (A.7). Dividing by $V(W,\rho,a,\lambda)^{\frac{1}{2}}V(W',\rho',a',\lambda')^{\frac{1}{2}}$, (A.7) can be written as

$$\{q\left[(\rho,a,\lambda),(\rho',a',\lambda')\right],P\left[(\rho,a,\lambda),(\rho',a',\lambda')\right]\}$$
$$=\arg\max_{q,P}\left[1-e^{-r\gamma[J(\rho,a+q,\lambda)-J(\rho,a,\lambda)-qP]}\right]^{\frac{1}{2}}\left[1-e^{-r\gamma[J(\rho',a'-q,\lambda')-J(\rho',a',\lambda')+qP]}\right]^{\frac{1}{2}},$$

s.t.

$$1-e^{-r\gamma[J(\rho,a+q,\lambda)-J(\rho,a,\lambda)-qP]}\geq 0$$
$$1-e^{-r\gamma[J(\rho',a'-q,\lambda')-J(\rho',a',\lambda')+qP]}\geq 0.$$

As can be seen, terms of trade are independent of wealth levels. Solving this problem is relatively straightforward: I set up the Lagrangian of this problem. Then using the first-order and Kuhn-Tucker conditions, the trade size $q\left[(\rho,a,\lambda),(\rho',a',\lambda')\right]$ solves the equation (8). And, the transaction price $P\left[(\rho,a,\lambda),(\rho',a',\lambda')\right]$ is given by the equation (9) if $J_2(\rho,a,\lambda)\neq J_2(\rho',a',\lambda')$; and $P=J_2(\rho,a,\lambda)$ if $J_2(\rho,a,\lambda)=J_2(\rho',a',\lambda')$. Substituting the transaction price into (A.12), I get

$$rJ(\rho,a,\lambda) = u(\rho,a) + \alpha \int_{-1}^{1} \frac{1-e^{-r\gamma[J(\rho',a,\lambda)-J(\rho,a,\lambda')]}}{r\gamma} dF(\rho')$$

$$+ \int_{-\infty}^{\infty}\int_{-1}^{1} \frac{1-e^{-\frac{r\gamma}{2}\{J(\rho,a+q[(\rho,a,\lambda),(\rho',a',\lambda')],\lambda)-J(\rho,a,\lambda)+J(\rho',a'-q[(\rho,a,\lambda),(\rho',a',\lambda')],\lambda')-J(\rho',a',\lambda')\}}}{r\gamma}$$

$$2\mu\lambda\frac{\lambda'}{\Lambda}\Phi(d\rho',da',d\lambda'), \quad \text{(A.13)}$$

subject to (8).

Equation (A.13) cannot be solved in closed-form. Consequently, following Gârleanu (2009), I use the linearization $\frac{1-e^{-r\gamma x}}{r\gamma} \approx x$ that ignores terms of order higher than 1 in $[J(\rho',a,\lambda) - J(\rho,a,\lambda)]$. The same approximation is also used by Biais (1993), Duffie et al. (2007), Vayanos and Weill (2008), Praz (2014) and Cujean and Praz (2015). Economic meaning of this approximation is that I assume investors are risk averse towards diffusion risks while they are risk neutral towards jump risks. The assumption does not suppress the impact of risk aversion as investors' preferences feature the fundamental risk-return trade-off associated with asset holdings. It only linearizes the preferences of investors over jumps in the continuation values created by trade or idiosyncratic shocks. The approximation yields the following lemma.

Lemma 3 *Fix parameters $\overline{\gamma}$, $\overline{\sigma}_D$ and $\overline{\sigma}_\eta$, and let $\sigma_D = \overline{\sigma}_D\sqrt{\overline{\gamma}/\gamma}$ and $\sigma_\eta = \overline{\sigma}_\eta\sqrt{\overline{\gamma}/\gamma}$. In any stationary equilibrium, investors' value functions solve the following HJB equation in the limit as γ goes to zero:*

$$rJ(\rho,a,\lambda) = am_D - \frac{1}{2}r\overline{\gamma}\left(a^2\overline{\sigma}_D^2 + 2\rho a\overline{\sigma}_D\overline{\sigma}_\eta\right) + \alpha\int_{-1}^{1}[J(\rho',a,\lambda) - J(\rho,a,\lambda)]dF(\rho')$$

$$+\int_0^1\int_{-\infty}^{\infty}\int_{-1}^{1}\mu\lambda\frac{\lambda'}{\Lambda}\left\{J(\rho, a + q\left[(\rho,a,\lambda),(\rho',a',\lambda')\right],\lambda) - J(\rho,a,\lambda)\right.$$

$$\left.+J(\rho', a' - q\left[(\rho,a,\lambda),(\rho',a',\lambda')\right],\lambda') - J(\rho',a',\lambda')\right\}\Phi(d\rho',da',d\lambda'), \quad \text{(A.14)}$$

subject to (8).

Ignoring the bars on γ, σ_D and σ_η, the problem is equivalent to the one with the reduced-form mean-variance utility flow.

Appendix B. Proofs

B.1 Proof of Theorem 1 and Proposition 2

After substituting the solution of Nash bargaining, the investors' problem is

$$rJ(\rho,a,\lambda) = am_D - \frac{1}{2}r\gamma\left(a^2\sigma_D^2 + 2\rho a\sigma_D\sigma_\eta\right) + \alpha\int_{-1}^{1}[J(\rho',a,\lambda) - J(\rho,a,\lambda)]dF(\rho')$$

$$+\int_0^1\int_{-\infty}^{\infty}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}\left[\max_q\left\{\frac{J(\rho,a+q,\lambda) - J(\rho,a,\lambda)}{2}\right.\right.$$

$$\left.\left.+\frac{J(\rho',a'-q,\lambda') - J(\rho',a',\lambda')}{2}\right\}\right]\Phi(d\rho',da',d\lambda').$$

Conjecture that

$$J(\rho,a,\lambda) = D(\lambda) + E(\lambda)\rho + F(\lambda)a + G(\lambda)a^2 + H(\lambda)\rho a + M(\lambda)\rho^2, \quad \text{(B.1)}$$

implying

$$J_2(\rho, a, \lambda) = F(\lambda) + 2G(\lambda)\, a + H(\lambda)\, \rho \tag{B.2}$$

and

$$J_{22}(\rho, a, \lambda) = 2G(\lambda). \tag{B.3}$$

Therefore, the value function can be written as

$$J(\rho, a, \lambda) = -G(\lambda)\, a^2 + J_2(\rho, a, \lambda)a + D(\lambda) + E(\lambda)\, \rho + M(\lambda)\, \rho^2. \tag{B.4}$$

$q[(\rho, a, \lambda), (\rho', a', \lambda')]$ is given by (8). Using the conjecture,

$$F(\lambda) + 2G(\lambda)\, a + 2G(\lambda)\, q + H(\lambda)\, \rho = F(\lambda') + 2G(\lambda')\, a' - 2G(\lambda')\, q + H(\lambda')\, \rho'.$$

Therefore,

$$q = \frac{J_2(\rho', a', \lambda') - J_2(\rho, a, \lambda)}{2\left(G(\lambda) + G(\lambda')\right)}.$$

Substituting back inside the conjectured marginal valuation, the post-trade marginal valuation is

$$J_2(\rho, a + q, \lambda) = J_2(\rho', a' - q, \lambda') = G(\lambda)\, \frac{J_2(\rho', a', \lambda')}{G(\lambda) + G(\lambda')} + G(\lambda')\, \frac{J_2(\rho, a, \lambda)}{G(\lambda) + G(\lambda')}. \tag{B.5}$$

$P[(\rho, a, \lambda), (\rho', a', \lambda')]$ is given by (9). Using the fact that $J(\rho, a, \lambda)$ is quadratic in a, a second-order Taylor expansion shows that:

$$J(\rho, a + q, \lambda) - J(\rho, a, \lambda) = J_2(\rho, a + q, \lambda)q - G(\lambda)\, q^2.$$

Then, Equation (9) implies

$$P = \frac{q}{2}\left(G\left(\lambda'\right) - G\left(\lambda\right)\right) + J_2(\rho, a+q, \lambda).$$

Hence, the terms of trade satisfy the system

$$q = \frac{J_2(\rho', a', \lambda') - J_2(\rho, a, \lambda)}{2\left(G\left(\lambda\right) + G\left(\lambda'\right)\right)}, \tag{B.6a}$$

$$P = \frac{q}{2}\left(G\left(\lambda'\right) - G\left(\lambda\right)\right) + G\left(\lambda\right)\frac{J_2(\rho', a', \lambda')}{G\left(\lambda\right) + G\left(\lambda'\right)} + G\left(\lambda'\right)\frac{J_2(\rho, a, \lambda)}{G\left(\lambda\right) + G\left(\lambda'\right)}. \tag{B.6b}$$

Using (B.5) and (B.6a), the implied trade surplus is

$$\begin{aligned}
&J(\rho, a+q, \lambda) - J(\rho, a, \lambda) + J(\rho', a'-q, \lambda') - J(\rho', a', \lambda') \\
&= -G\left(\lambda\right)\left(2aq + q^2\right) + J_2(\rho, a+q, \lambda)\left(a+q\right) - J_2(\rho, a, \lambda)a \\
&- G\left(\lambda'\right)\left(-2a'q + q^2\right) + J_2(\rho', a'-q, \lambda')\left(a'-q\right) - J_2(\rho', a', \lambda')a' \\
&= -\frac{\left(J_2(\rho', a', \lambda') - J_2(\rho, a, \lambda)\right)^2}{4\left(G\left(\lambda\right) + G\left(\lambda'\right)\right)}.
\end{aligned}$$

Rewrite the investors' problem by substituting the trade surplus implied by the Nash bargaining solution:

$$\begin{aligned}
rJ(\rho, a, \lambda) = am_D - \frac{1}{2}r\gamma\left(a^2\sigma_D^2 + 2\rho a\sigma_D\sigma_\eta\right) + \alpha\int_{-1}^{1}[J(\rho', a, \lambda) - J(\rho, a, \lambda)]dF(\rho') \\
+\int_0^1\int_{-\infty}^{\infty}\int_{-1}^{1} 2\mu\lambda\frac{\lambda'}{\Lambda}\left\{-\frac{\left(J_2(\rho', a', \lambda') - J_2(\rho, a, \lambda)\right)^2}{8\left(G\left(\lambda\right) + G\left(\lambda'\right)\right)}\right\}\Phi(d\rho', da', d\lambda').
\end{aligned} \tag{B.7}$$

Therefore, my conjectured value function is verified after substituting the Nash bargaining solution. The marginal valuation satisfies the flow Bellman equation:

$$rJ_2(\rho,a,\lambda) = m_D - r\gamma\left(a\sigma_D^2 + \rho\sigma_D\sigma_\eta\right) + \alpha\int_{-1}^{1}[J_2(\rho',a,\lambda) - J_2(\rho,a,\lambda)]dF(\rho')$$
$$+\int_0^1\int_{-\infty}^{\infty}\int_{-1}^{1} 2\mu\lambda\frac{\lambda'}{\Lambda}\left\{\frac{J_2(\rho',a',\lambda') - J_2(\rho,a,\lambda)}{4\left(G(\lambda)+G(\lambda')\right)}2G(\lambda)\right\}\Phi(d\rho',da',d\lambda'). \tag{B.8}$$

Taking all terms which contain $J_2(\rho,a,\lambda)$ to the LHS,

$$\left(r+\alpha+\int_0^1 \mu\lambda\frac{\lambda'}{\Lambda}\frac{G(\lambda)}{G(\lambda)+G(\lambda')}d\Psi(\lambda')\right)J_2(\rho,a,\lambda) = m_D - r\gamma\left(a\sigma_D^2+\rho\sigma_D\sigma_\eta\right)$$
$$+\alpha\int_{-1}^{1} J_2(\rho',a,\lambda)dF(\rho') + \int_0^1\int_{-\infty}^{\infty}\int_{-1}^{1}\mu\lambda\frac{\lambda'}{\Lambda}\frac{G(\lambda)}{G(\lambda)+G(\lambda')}J_2(\rho',a',\lambda')\Phi(d\rho',da',d\lambda').$$

Substitute the conjectured marginal valuation and match coefficients:

$$\left(\alpha+\widetilde{r}(\lambda)\right)\left(F(\lambda)+2G(\lambda)a+H(\lambda)\rho\right)$$
$$= m_D - r\gamma\left(a\sigma_D^2+\rho\sigma_D\sigma_\eta\right) + \alpha\int_{-1}^{1}\left[F(\lambda)+2G(\lambda)a+H(\lambda)\rho'\right]dF(\rho') + \left(\widetilde{r}(\lambda)-r\right)\overline{J}_2(\lambda),$$

where

$$\widetilde{r}(\lambda) \equiv r + \int_0^1 \mu\lambda\frac{\lambda'}{\Lambda}\frac{G(\lambda)}{G(\lambda)+G(\lambda')}d\Psi(\lambda'),$$
$$\overline{J}_2(\lambda) \equiv \frac{\int_0^1\int_{-\infty}^{\infty}\int_{-1}^{1}\mu\lambda\frac{\lambda'}{\Lambda}\frac{G(\lambda)}{G(\lambda)+G(\lambda')}J_2(\rho',a',\lambda')\Phi(d\rho',da',d\lambda')}{\widetilde{r}(\lambda)-r},$$

where $\widetilde{r}(\lambda)$ is the effective discount rate of an investor with λ. Equivalently,

$$\begin{aligned}&(\alpha+\widetilde{r}(\lambda))\left(F(\lambda)+2G(\lambda)a+H(\lambda)\rho\right)\\&\quad=m_D-r\gamma\left(a\sigma_D^2+\rho\sigma_D\sigma_\eta\right)+\alpha\left(F(\lambda)+2G(\lambda)a+H(\lambda)\overline{\rho}\right)+\left(\widetilde{r}(\lambda)-r\right)\overline{J}_2(\lambda).\end{aligned}$$

Then, undetermined coefficients solve the system:

$$\widetilde{r}(\lambda)F(\lambda)=m_D+\alpha H(\lambda)\overline{\rho}+\left(\widetilde{r}(\lambda)-r\right)\overline{J}_2(\lambda), \tag{B.9}$$

$$\widetilde{r}(\lambda)2G(\lambda)=-r\gamma\sigma_D^2, \tag{B.10}$$

$$(\alpha+\widetilde{r}(\lambda))H(\lambda)=-r\gamma\sigma_D\sigma_\eta. \tag{B.11}$$

Using the resulting G from the matched coefficients, the definition of $\widetilde{r}(\lambda)$ implies

$$\widetilde{r}(\lambda)=r+\int_0^1\mu\lambda\frac{\lambda'}{\Lambda}\frac{\frac{-r\gamma\sigma_D^2}{2\widetilde{r}(\lambda)}}{\frac{-r\gamma\sigma_D^2}{2\widetilde{r}(\lambda)}+\frac{-r\gamma\sigma_D^2}{2\widetilde{r}(\lambda')}}d\Psi(\lambda').$$

Then, $\widetilde{r}(\lambda)$ satisfies the recursive functional equation:

$$\widetilde{r}(\lambda)=r+\int_0^1\mu\lambda\frac{\lambda'}{\overline{\lambda}}\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')}d\Psi(\lambda'). \tag{B.13}$$

Using the matched coefficients,

$$J_2(\rho,a,\lambda)=\frac{m_D-r\gamma\sigma_D^2a-r\gamma\sigma_D\sigma_\eta\frac{\widetilde{r}(\lambda)\rho+\alpha\overline{\rho}}{\widetilde{r}(\lambda)+\alpha}+\left(\widetilde{r}(\lambda)-r\right)\overline{J}_2(\lambda)}{\widetilde{r}(\lambda)}, \tag{B.14}$$

where

$$\overline{J}_2(\lambda) = \frac{\int_0^1 \int_{-\infty}^{\infty} \int_{-1}^{1} \mu\lambda \frac{\lambda'}{\Lambda} \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')} J_2(\rho', a', \lambda') \Phi(d\rho', da', d\lambda')}{\widetilde{r}(\lambda) - r}. \tag{B.15}$$

To complete the proof of Theorem 1, I need to show that $\overline{J}_2(\lambda) = \frac{u_2(\overline{\rho}, A)}{r}$. Using (B.14):

$$\overline{J}_2(\lambda) = \frac{\int_0^1 \int_{-\infty}^{\infty} \int_{-1}^{1} \mu\lambda \frac{\lambda'}{\Lambda} \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')} \left[\frac{m_D - r\gamma\sigma_D^2 a' - r\gamma\sigma_D\sigma_\eta \frac{\widetilde{r}(\lambda')\rho' + \alpha\overline{\rho}}{\widetilde{r}(\lambda')+\alpha} + (\widetilde{r}(\lambda') - r)\overline{J}_2(\lambda')}{\widetilde{r}(\lambda')} \right] \Phi(d\rho', da', d\lambda')}{\widetilde{r}(\lambda) - r}.$$

After cancellations, and using the fact that measure of specialists is independent of idiosyncratic correlation shocks,

$$(\widetilde{r}(\lambda) - r)\overline{J}_2(\lambda) = \int_0^1 \mu\lambda \frac{\lambda'}{\Lambda} \frac{1}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')} \left(m_D - r\gamma\sigma_D\sigma_\eta\overline{\rho} - r\gamma\sigma_D^2 \mathbb{E}_\phi [a' \mid \lambda'] + (\widetilde{r}(\lambda') - r)\overline{J}_2(\lambda') \right) d\Psi(\lambda'). \tag{B.16}$$

This equation reveals that the expected contribution of the market to an investor's post-trade marginal valuation depends on the mean of equilibrium holdings $E_\phi [a' \mid \lambda']$ conditional on measure of trading specialists. It will be determined when I derive the first moment of equilibrium distribution. Thus, the proof of Theorem 1 will be complete after the proof of Proposition 2. The following lemma constitutes the starting point of the proof of Proposition 2.

Lemma 4 *Given $\overline{J}_2(\lambda)$, the conditional pdf $\phi_{\rho\lambda}(a)$ of asset holdings satisfies the system*

$$(\alpha + 2\mu\lambda)\,\phi_{\rho,\lambda}(a) = \alpha \int_{-1}^{1} \phi_{\rho',\lambda}(a)\, dF(\rho')$$

$$+ \int_{0}^{1}\int_{-1}^{1}\int_{-\infty}^{\infty} 2\mu\lambda \frac{\lambda'}{\Lambda}\left(1 + \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)\phi_{\rho,\lambda}(a')$$

$$\phi_{\rho',\lambda'}\left(a\left(1 + \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right) - a' - \widetilde{m}_D(\lambda,\lambda') + \widetilde{C}\left[(\rho,\lambda),(\rho',\lambda')\right] - \widetilde{J}(\lambda,\lambda')\right) da'\, dF(\rho')\, d\Psi(\lambda'),$$

where

$$\widetilde{m}_D(\lambda,\lambda') \equiv \frac{\widetilde{r}(\lambda') - \widetilde{r}(\lambda)}{r\gamma\sigma_D^2 \widetilde{r}(\lambda)} m_D,$$

$$\widetilde{C}\left[(\rho,\lambda),(\rho',\lambda')\right] \equiv \frac{\sigma_\eta}{\sigma_D}\left(\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\frac{\widetilde{r}(\lambda)\rho + \alpha\overline{\rho}}{\widetilde{r}(\lambda) + \alpha} - \frac{\widetilde{r}(\lambda')\rho' + \alpha\overline{\rho}}{\widetilde{r}(\lambda') + \alpha}\right),$$

$$\widetilde{J}(\lambda,\lambda') \equiv \frac{\widetilde{r}(\lambda')}{r\gamma\sigma_D^2 \widetilde{r}(\lambda)}\left(\widetilde{r}(\lambda) - r\right)\overline{J}_2(\lambda) - \frac{1}{r\gamma\sigma_D^2}\left(\widetilde{r}(\lambda') - r\right)\overline{J}_2(\lambda').$$

With further simplification,

$$(\alpha + 2\mu\lambda)\,\phi_{\rho,\lambda}(a) = \alpha \int_{-1}^{1} \phi_{\rho',\lambda}(a)\, dF(\rho')$$

$$+ \int_{0}^{1}\int_{-1}^{1}\int_{-\infty}^{\infty} 2\mu\lambda \frac{\lambda'}{\Lambda}\left(1 + \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)\phi_{\rho,\lambda}(a')$$

$$\phi_{\rho',\lambda'}\left(a\left(1 + \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right) - a' + \overline{C}\left[(\rho,\lambda),(\rho',\lambda')\right]\right) da'\, dF(\rho')\, d\Psi(\lambda'),$$

where

$$\overline{C}\left[(\rho,\lambda),(\rho',\lambda')\right] \equiv -\widetilde{m}_D(\lambda,\lambda') + \widetilde{C}\left[(\rho,\lambda),(\rho',\lambda')\right] - \widetilde{J}(\lambda,\lambda').$$

Taking the Fourier transform of the steady-state condition above, the first equation of the proposition 2 is proven. The second equation comes from the fact that $\phi_{\rho,\lambda}(a)$ is a pdf. And, the third equation is implied by market clearing. When I derive $\widetilde{C}[(\rho,\lambda),(\rho',\lambda')]$, the proof will be complete.

The first derivative of the Fourier transform evaluated at $z=0$ is

$$\begin{aligned}
(\alpha+2\mu\lambda)\,\widehat{\phi}'_{\rho,\lambda}(0) &= \alpha\int_{-1}^{1}\widehat{\phi}'_{\rho',\lambda}(0)\,dF(\rho') \\
&+\int_{0}^{1}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}\frac{1}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\widehat{\phi}'_{\rho,\lambda}(0)\,dF(\rho')\,d\Psi(\lambda') \\
&+\int_{0}^{1}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}i2\pi\overline{C}[(\rho,\lambda),(\rho',\lambda')]\frac{1}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}dF(\rho')\,d\Psi(\lambda') \\
&+\int_{0}^{1}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}\frac{1}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\widehat{\phi}'_{\rho',\lambda'}(0)\,dF(\rho')\,d\Psi(\lambda').
\end{aligned}$$

Therefore, the first moments satisfy

$$\begin{aligned}
(\alpha+2\mu\lambda)\,\mathbb{E}_{\phi}[a\mid\rho,\lambda] &= \alpha\int_{-1}^{1}\mathbb{E}_{\phi}[a\mid\rho',\lambda]\,dF(\rho') \\
&+\int_{0}^{1}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}\frac{1}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\mathbb{E}_{\phi}[a\mid\rho,\lambda]\,dF(\rho')\,d\Psi(\lambda') \\
&-\int_{0}^{1}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}\overline{C}[(\rho,\lambda),(\rho',\lambda')]\frac{1}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}dF(\rho')\,d\Psi(\lambda') \\
&+\int_{0}^{1}\int_{-1}^{1}2\mu\lambda\frac{\lambda'}{\Lambda}\frac{1}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\mathbb{E}_{\phi}[a\mid\rho',\lambda']\,dF(\rho')\,d\Psi(\lambda'),
\end{aligned}$$

$$
\begin{aligned}
(\alpha + 2\mu\lambda)\,\mathbb{E}_\phi [a \mid \rho, \lambda] &= \alpha \mathbb{E}_\phi [a \mid \lambda] + \mathbb{E}_\phi [a \mid \rho, \lambda]\, 2\left(r + \mu\lambda - \widetilde{r}(\lambda)\right) \\
&\quad - \int_0^1 2\mu\lambda \frac{\lambda'}{\Lambda} \overline{C}\left[(\rho, \lambda), (\rho', \lambda')\right] \frac{1}{1 + \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} d\Psi(\lambda') \\
&\quad + \int_0^1 2\mu\lambda \frac{\lambda'}{\Lambda} \frac{1}{1 + \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} \mathbb{E}_\phi [a \mid \lambda']\, d\Psi(\lambda'),
\end{aligned}
$$

$$
\begin{aligned}
(\alpha + 2(\widetilde{r}(\lambda) - r))\,\mathbb{E}_\phi [a \mid \rho, \lambda] &= \alpha \mathbb{E}_\phi [a \mid \lambda] \\
&\quad - \int_0^1 2\mu\lambda \frac{\lambda'}{\Lambda} \overline{C}\left[(\rho, \lambda), (\rho', \lambda')\right] \frac{1}{1 + \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} d\Psi(\lambda') \\
&\quad + \int_0^1 2\mu\lambda \frac{\lambda'}{\Lambda} \frac{1}{1 + \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} \mathbb{E}_\phi [a \mid \lambda']\, d\Psi(\lambda'),
\end{aligned}
$$

where the second term is

$$
\begin{aligned}
&\int_0^1 2\mu\lambda \frac{\lambda'}{\Lambda} \overline{C}\left[(\rho, \lambda), (\rho', \lambda')\right] \frac{1}{1 + \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} d\Psi(\lambda') \\
&\quad = \int_0^1 2\mu\lambda \frac{\lambda'}{\Lambda} \frac{1}{r\gamma\sigma_D^2} \left[-\left(\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)} - 1 \right) m_D + r\gamma\sigma_D\sigma_\eta \left(\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)} \frac{\widetilde{r}(\lambda)\rho + \alpha\overline{\rho}}{\widetilde{r}(\lambda) + \alpha} - \overline{\rho} \right) \right. \\
&\qquad \left. - \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)} (\widetilde{r}(\lambda) - r)\, \overline{J}_2(\lambda) + (\widetilde{r}(\lambda') - r)\, \overline{J}_2(\lambda') \right] \frac{1}{1 + \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} d\Psi(\lambda').
\end{aligned}
$$

Take expectation over ρ, and substitute out $\overline{C}\left[(\rho, \lambda), (\rho', \lambda')\right]$:

$$
\begin{aligned}
(\widetilde{r}(\lambda) - r)\,\mathbb{E}_\phi [a \mid \lambda] &= -\int_0^1 \mu\lambda \frac{\lambda'}{\Lambda} \frac{1}{r\gamma\sigma_D^2} \left[-\left(\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)} - 1 \right) (m_D - r\gamma\sigma_D\sigma_\eta\overline{\rho}) \right. \\
&\quad \left. - \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)} (\widetilde{r}(\lambda) - r)\, \overline{J}_2(\lambda) + (\widetilde{r}(\lambda') - r)\, \overline{J}_2(\lambda') \right] \frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')} d\Psi(\lambda') \\
&\quad + \int_0^1 \mu\lambda \frac{\lambda'}{\Lambda} \frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')} \mathbb{E}_\phi [a \mid \lambda']\, d\Psi(\lambda').
\end{aligned}
$$

And note that the equation (B.16) also connects $\overline{J}_2(\lambda')$ and $E_\phi[a \mid \lambda']$ as a result of optimality:

$$(\widetilde{r}(\lambda) - r)\,\overline{J}_2(\lambda) = (m_D - r\gamma\sigma_D\sigma_\eta\overline{\rho})\left(\frac{r+\mu\lambda}{\widetilde{r}(\lambda)} - 1\right) + \int_0^1 \mu\lambda\frac{\lambda'}{\Lambda}\frac{1}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')}\left(-r\gamma\sigma_D^2\mathbb{E}_\phi[a' \mid \lambda'] + (\widetilde{r}(\lambda') - r)\,\overline{J}_2(\lambda')\right)d\Psi(\lambda').$$

Thus, the last two equations combined with the market-clearing condition

$$\int_0^1 \mathbb{E}_\phi[a' \mid \lambda']\,d\Psi(\lambda') = A$$

pin down $E_\phi[a \mid \lambda]$ and $\overline{J}_2(\lambda)$ for all $\lambda \in$supp(Ψ). Since λ takes values on a finite set, it is easy to verify that the conditions imply a non-singular linear system with the unique solution:

$$\begin{aligned}\mathbb{E}_\phi[a \mid \lambda] &= A,\\ \overline{J}_2(\lambda) &= \frac{m_D}{r} - \gamma\sigma_D\sigma_\eta\overline{\rho} - \gamma\sigma_D^2 A.\end{aligned}$$

This completes the proof of Theorem 1. Using this solution,

$$\widetilde{J}(\lambda,\lambda') = -\frac{\widetilde{r}(\lambda') - \widetilde{r}(\lambda)}{\gamma\sigma_D^2\widetilde{r}(\lambda)}\left(\frac{m_D}{r} - \gamma\sigma_D\sigma_\eta\overline{\rho} - \gamma\sigma_D^2 A\right),$$

which implies

$$\overline{C}[(\rho,\lambda),(\rho',\lambda')] = \widetilde{r}(\lambda')\frac{\sigma_\eta}{\sigma_D}\left(\frac{\rho-\overline{\rho}}{\widetilde{r}(\lambda)+\alpha} - \frac{\rho'-\overline{\rho}}{\widetilde{r}(\lambda')+\alpha}\right) - \left(\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)} - 1\right)A,$$

and the proof Proposition 2 is also complete.

Proposition 1 can be derived as a by-product of the steps in this proof. More precisely, (17) is derived by substituting $\overline{J}_2(\lambda)$ into (B.14). Using the resulting formula for marginal valuation and (B.10), equations (B.6a) and (B.6b) imply (18) and (19), respectively.

Using the marginal valuation in Proposition 1, application of the method of undetermined coefficients to (B.7) pins down all the coefficients in (B.1):

$$(r+\alpha) M(\lambda) = \frac{r\gamma\sigma_\eta^2}{2\left(\widetilde{r}(\lambda)+\alpha\right)^2}\widetilde{r}(\lambda)\left(\widetilde{r}(\lambda)-r\right),$$

$$(r+\alpha) E(\lambda) = H(\lambda)\int_0^1 2\mu\lambda\frac{\lambda'}{\Lambda}\frac{F(\lambda')+2G(\lambda')A+H(\lambda')\overline{\rho}-F(\lambda)}{4\left(G(\lambda)+G(\lambda')\right)}d\Psi(\lambda'),$$

$$\begin{aligned} rD(\lambda) = {} & \alpha\left(E(\lambda)\overline{\rho}+M(\lambda)\overline{\rho^2}\right) \\ & + \int_0^1\int_{-\infty}^{\infty}\int_{-1}^{1} 2\mu\lambda\frac{\lambda'}{\Lambda}\left\{-\frac{\left[F(\lambda')+2G(\lambda')a'+H(\lambda')\rho'-F(\lambda)\right]^2}{8\left(G(\lambda)+G(\lambda')\right)}\right\}\Phi(d\rho',da',d\lambda'). \end{aligned}$$

Therefore, the value function is available in closed form up to effective discount rates $\widetilde{r}(\lambda)$.

B.2 Proof of Lemma 4

Assuming $\Phi_\lambda(\rho,a)$ is the joint cdf of correlations and asset holdings conditional on search intensity, rearrangement of the equation (7) yields

$$
\begin{aligned}
0 = &-\alpha\Phi_{\lambda^*}(\rho^*, a^*) + \alpha \int_{-\infty}^{a^*}\int_{-1}^{1} \Phi_{\lambda^*}(d\rho, da) F(\rho^*) \\
&- \frac{2\mu\lambda^*}{\Lambda} \int_{-\infty}^{a^*}\int_{-1}^{\rho^*} \left[\int_0^1 \int_{-\infty}^{\infty}\int_{-1}^{1} \lambda' \mathbb{I}_{\{q[(\rho,a,\lambda^*),(\rho',a',\lambda')]\geq a^*-a\}} \Phi_{\lambda'}(d\rho', da') d\Psi\left(\lambda'\right) \right] \Phi_{\lambda^*}(d\rho, da) \\
&+ \frac{2\mu\lambda^*}{\Lambda} \int_{a^*}^{\infty}\int_{-1}^{\rho^*} \left[\int_0^1 \int_{-\infty}^{\infty}\int_{-1}^{1} \lambda' \mathbb{I}_{\{q[(\rho,a,\lambda^*),(\rho',a',\lambda')]< a^*-a\}} \Phi_{\lambda'}(d\rho', da') d\Psi\left(\lambda'\right) \right] \Phi_{\lambda^*}(d\rho, da)
\end{aligned}
$$

for all $\lambda^* \in \text{supp}(\Psi)$. For simplicity, I assume that the distribution of correlations and the equilibrium conditional distribution of asset holdings have densities. This assumption is actually never used but simplifies the presentation of the results. I write the above condition in terms of conditional pdfs, by letting $\phi_{\rho,\lambda}(a)$ denote the conditional pdf of asset holdings by investors with correlation ρ and search intensity λ:

$$
\begin{aligned}
0 = &-\alpha \int_{-1}^{\rho^*}\int_{-\infty}^{a^*} \phi_{\rho,\lambda^*}(a) da dF(\rho) + \alpha \int_{-1}^{1}\int_{-\infty}^{a^*} \phi_{\rho,\lambda^*}(a) da dF(\rho) F(\rho^*) \\
&- \frac{2\mu\lambda^*}{\Lambda} \int_{-1}^{\rho^*}\int_{-\infty}^{a^*} \left[\int_0^1 \int_{-1}^{1}\int_{-\infty}^{\infty} \lambda' \mathbb{I}_{\{q[(\rho,a,\lambda^*),(\rho',a',\lambda')]\geq a^*-a\}} \phi_{\rho',\lambda'}(a') da' dF(\rho') d\Psi\left(\lambda'\right) \right] \phi_{\rho,\lambda^*}(a) da dF(\rho) \\
&+ \frac{2\mu\lambda^*}{\Lambda} \int_{-1}^{\rho^*}\int_{a^*}^{\infty} \left[\int_0^1 \int_{-1}^{1}\int_{-\infty}^{\infty} \lambda' \mathbb{I}_{\{q[(\rho,a,\lambda^*),(\rho',a',\lambda')]< a^*-a\}} \phi_{\rho',\lambda'}(a') da' dF(\rho') d\Psi\left(\lambda'\right) \right] \phi_{\rho,\lambda^*}(a) da dF(\rho).
\end{aligned}
$$

Using the expression for trade sizes implied by (B.6a), I can get rid of indicator functions inside the integrals, using appropriate bounds:

$$0=-\alpha\int_{-1}^{\rho^*}\int_{-\infty}^{a^*}\phi_{\rho,\lambda^*}(a)dadF(\rho)+\alpha F(\rho^*)\int_{-1}^{1}\int_{-\infty}^{a^*}\phi_{\rho,\lambda^*}(a)dadF(\rho)$$

$$-\frac{2\mu\lambda^*}{\Lambda}\int_{-1}^{\rho^*}\int_{-\infty}^{a^*}\left[\int_0^1\int_{-1}^1\int_{\xi[(\rho,a,\lambda^*),(\rho',a',\lambda')]}^{\infty}\lambda'\phi_{\rho',\lambda'}(a')da'dF(\rho')d\Psi(\lambda')\right]\phi_{\rho,\lambda^*}(a)dadF(\rho)$$

$$+\frac{2\mu\lambda^*}{\Lambda}\int_{-1}^{\rho^*}\int_{a^*}^{\infty}\left[\int_0^1\int_{-1}^1\int_{-\infty}^{\xi[(\rho,a,\lambda^*),(\rho',a',\lambda')]}\lambda'\phi_{\rho',\lambda'}(a')da'dF(\rho')d\Psi(\lambda')\right]\phi_{\rho,\lambda^*}(a)dadF(\rho),$$

where

$$\xi\left[(\rho,a,\lambda),(\rho',a',\lambda')\right]=a\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)-a'-\widetilde{m}_D(\lambda,\lambda')+\widetilde{C}\left[(\rho,\lambda),(\rho',\lambda')\right]-\widetilde{J}(\lambda,\lambda'),$$

$$\widetilde{m}_D(\lambda,\lambda')\equiv\frac{\widetilde{r}(\lambda')-\widetilde{r}(\lambda)}{r\gamma\sigma_D^2\widetilde{r}(\lambda)}m_D,$$

$$\widetilde{C}\left[(\rho,\lambda),(\rho',\lambda')\right]\equiv\frac{\sigma_\eta}{\sigma_D}\left(\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\frac{\widetilde{r}(\lambda)\rho+\alpha\overline{\rho}}{\widetilde{r}(\lambda)+\alpha}-\frac{\widetilde{r}(\lambda')\rho'+\alpha\overline{\rho}}{\widetilde{r}(\lambda')+\alpha}\right),$$

$$\widetilde{J}(\lambda,\lambda')\equiv\frac{\widetilde{r}(\lambda')}{r\gamma\sigma_D^2\widetilde{r}(\lambda)}\left(\widetilde{r}(\lambda)-r\right)\overline{J}_2(\lambda)-\frac{1}{r\gamma\sigma_D^2}\left(\widetilde{r}(\lambda')-r\right)\overline{J}_2(\lambda').$$

Since this equality holds for any (ρ^*,a^*,λ^*), one can take derivative of the both sides with respect to ρ^* using Leibniz rule whenever necessary:

$$0=-\alpha f(\rho^*)\int_{-\infty}^{a^*}\phi_{\rho^*,\lambda^*}(a)da+\alpha f(\rho^*)\int_{-1}^{1}\int_{-\infty}^{a^*}\phi_{\rho,\lambda^*}(a)dadF(\rho)$$

$$-\frac{2\mu\lambda^*}{\Lambda}f(\rho^*)\int_{-\infty}^{a^*}\left[\int_0^1\int_{-1}^1\int_{\xi[(\rho^*,a,\lambda^*),(\rho',a',\lambda')]}^{\infty}\lambda'\phi_{\rho',\lambda'}(a')da'dF(\rho')d\Psi(\lambda')\right]\phi_{\rho^*,\lambda^*}(a)da$$

$$+\frac{2\mu\lambda^*}{\Lambda}f(\rho^*)\int_{a^*}^{\infty}\left[\int_0^1\int_{-1}^1\int_{-\infty}^{\xi[(\rho^*,a,\lambda^*),(\rho',a',\lambda')]}\lambda'\phi_{\rho',\lambda'}(a')da'dF(\rho')d\Psi(\lambda')\right]\phi_{\rho^*,\lambda^*}(a)da.$$

After cancellations,

$$0=-\alpha\int_{-\infty}^{a^*}\phi_{\rho^*,\lambda^*}(a)da+\alpha\int_{-1}^{1}\int_{-\infty}^{a^*}\phi_{\rho,\lambda^*}(a)dadF(\rho)$$
$$-\frac{2\mu\lambda^*}{\Lambda}\int_{-\infty}^{a^*}\left[\int_0^1\int_{-1}^1\int_{\xi[(\rho^*,a,\lambda^*),(\rho',a',\lambda')]}^{\infty}\lambda'\phi_{\rho',\lambda'}(a')da'dF(\rho')d\Psi(\lambda')\right]\phi_{\rho^*,\lambda^*}(a)da$$
$$+\frac{2\mu\lambda^*}{\Lambda}\int_{a^*}^{\infty}\left[\int_0^1\int_{-1}^1\int_{-\infty}^{\xi[(\rho^*,a,\lambda^*),(\rho',a',\lambda')]}\lambda'\phi_{\rho',\lambda'}(a')da'dF(\rho')d\Psi(\lambda')\right]\phi_{\rho^*,\lambda^*}(a)da.$$

Similarly, take derivative with respect to a^* using Leibniz rule whenever necessary:

$$0=-\alpha\phi_{\rho^*,\lambda^*}(a^*)+\alpha\int_{-1}^{1}\phi_{\rho,\lambda^*}(a^*)dF(\rho)$$
$$-\frac{2\mu\lambda^*}{\Lambda}\int_{-\infty}^{a^*}\left[-\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)\int_0^1\int_{-1}^1\lambda'\phi_{\rho',\lambda'}(\xi\left[(\rho^*,a^*,\lambda^*),(\rho',a',\lambda')\right])dF(\rho')d\Psi(\lambda')\right]\phi_{\rho^*,\lambda^*}(a)da$$
$$-\frac{2\mu\lambda^*}{\Lambda}\int_{-\infty}^{a^*}\left[\int_0^1\int_{-1}^1\int_{\xi[(\rho^*,a^*,\lambda^*),(\rho',a',\lambda')]}^{\infty}\lambda'\phi_{\rho',\lambda'}(a')da'dF(\rho')d\Psi(\lambda')\right]\phi_{\rho^*,\lambda^*}(a^*)$$
$$+\frac{2\mu\lambda^*}{\Lambda}\int_{a^*}^{\infty}\left[\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)\int_0^1\int_{-1}^1\lambda'\phi_{\rho',\lambda'}(\xi\left[(\rho^*,a^*,\lambda^*),(\rho',a',\lambda')\right])dF(\rho')d\Psi(\lambda')\right]\phi_{\rho^*,\lambda^*}(a)da$$
$$-\frac{2\mu\lambda^*}{\Lambda}\left[\int_0^1\int_{-1}^1\int_{-\infty}^{\xi[(\rho^*,a^*,\lambda^*),(\rho',a',\lambda')]}\lambda'\phi_{\rho',\lambda'}(a')da'dF(\rho')d\Psi(\lambda')\right]\phi_{\rho^*,\lambda^*}(a^*).$$

After simplification, the Lemma is derived.

B.3 Proof of Lemma 1

Restate the equation (14):

$$\widetilde{r}(\lambda) = r + \int_0^1 \mu\lambda \frac{\lambda'}{\Lambda} \frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')} d\Psi(\lambda'),$$

where $\widetilde{r}(\lambda) > 0$ for all $\lambda \in$supp(Ψ) from the strict concavity of the value function. The functional equation, in turn, implies that $\widetilde{r}(\lambda) > r$ for all $\lambda \in$supp(Ψ). First, let's establish the existence and uniqueness of the solution of this functional equation. Rewrite:

$$\widetilde{r}(\lambda) = r + \int_0^1 \mu\lambda \frac{\lambda'}{\Lambda} d\Psi(\lambda') - \widetilde{r}(\lambda) \int_0^1 \mu\lambda \frac{\lambda'}{\Lambda} \frac{1}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')} d\Psi(\lambda').$$

Rearrangement yields an alternative representation of the functional equation:

$$\widetilde{r}(\lambda) = \frac{r + \mu\lambda}{1 + \int_0^1 \mu\lambda \frac{\lambda'}{\Lambda} \frac{1}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')} d\Psi(\lambda')}.$$

Since I assume a finite support, let supp$(\Psi) = \{\lambda_1, \lambda_2, ..., \lambda_N\}$ with ψ_n denoting the fraction of investors with λ_n for all $n \in \{1, 2, ..., N\}$. And let $\widetilde{r}_n = \widetilde{r}(\lambda_n)$ for all $n \in \{1, 2, ..., N\}$. Define the mapping $T : [0, \infty)^N \to [0, \infty)^N$ such that

$$T\widetilde{r}_n = \max \left\{ r, \frac{r + \mu\lambda_n}{1 + \sum_{k=1}^{N} \mu\lambda_n \frac{\lambda_k}{\Lambda} \frac{1}{\widetilde{r}_n + \widetilde{r}_k} \psi_k} \right\}.$$

$[0, \infty)^N$ with the usual sup norm constitutes a real Banach space. And, the set $[0, \infty)^N$ is a strongly minihedral cone itself (see Krasnosel'skiĭ, 1964). Thus, the solution of the functional equation is a non-zero fixed point of T on a strongly minihedral cone. Theorem

4.1 of Krasnosel'skiĭ (1964) shows that every monotone mapping on a strongly minihedral cone has at least one non-zero fixed point. It is easy to verify the monotonicity of T, i.e. $\widetilde{r}^A, \widetilde{r}^B \in [0,\infty)^N$ and $\widetilde{r}^A \leq \widetilde{r}^B$ imply $T\widetilde{r}^A \leq T\widetilde{r}^B$. Hence, the existence of the solution of the functional equation is established.

To show the uniqueness, I follow Theorem 6.3 of Krasnosel'skiĭ (1964), which states that every u_0-concave and monotone mapping on a cone has at most one non-zero fixed point. Therefore, it suffices to show that T is u_0-concave. By the definition of u_0-concavity, T is u_0-concave if there exists a non-zero element $u_0 \in [0,\infty)^N$ such that for an arbitrary non-zero $\widetilde{r} \in [0,\infty)^N$ there exist $b_l, b_u \in \mathbb{R}_{++}$ such that

$$b_l u_0 \leq T\widetilde{r} \leq b_u u_0,$$

and if for every $t_0 \in (0,1)$ there exists $\eta(t_0) \in \mathbb{R}_{++}$ such that

$$T(t_0\widetilde{r}) \geq (1+\eta(t_0))\, t_0 T\widetilde{r}.$$

It can be easily verified from the definition of T that these conditions are satisfied for $u_0 = (r+\mu, ..., r+\mu)$, $b_l = r(r+\mu)^{-1}$, $b_u = 1$, and $\eta(t_0) = (1-t_0)\left(t_0 + \frac{\mu}{2r\Lambda}\right)^{-1}$. Hence, the uniqueness of the solution of the functional equation is established as well.

The effective discount rate function is strictly increasing if $\widetilde{r}(\lambda') > \widetilde{r}(\lambda)$ for all $\lambda \in$supp(Ψ) and for all $\lambda' \in$supp(Ψ) with $\lambda' > \lambda$. To obtain a contradiction, suppose there exists $\lambda, \lambda' \in$supp(Ψ) with $\lambda' > \lambda$, and $\widetilde{r}(\lambda') \leq \widetilde{r}(\lambda)$. The equation (14) implies that $\widetilde{r}(\lambda')$ and

$\widetilde{r}(\lambda)$ satisfy the following equations respectively:

$$\widetilde{r}(\lambda') = r + \frac{\mu\lambda'}{\Lambda}\int_0^1 \frac{\lambda''\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda') + \widetilde{r}(\lambda'')} d\Psi(\lambda'')$$
$$\widetilde{r}(\lambda) = r + \frac{\mu\lambda}{\Lambda}\int_0^1 \frac{\lambda''\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda'')} d\Psi(\lambda'').$$

As $\lambda' > \lambda$ and $\widetilde{r}(\lambda') \leq \widetilde{r}(\lambda)$, the RHS of the second equation is lower than the RHS of the first equation, which implies that $\widetilde{r}(\lambda') > \widetilde{r}(\lambda)$; and we obtain the desired contradiction. Hence, the effective discount rate function is strictly increasing.

To show the strict concavity of effective discount rate function, I use the following definition of strict concavity for functions defined on a finite domain, adapted from Yüceer (2002).

Definition 2 Let $S \subset \mathbb{R}$ be a discrete one-dimentional space. A function $f : S \to \mathbb{R}$ is strictly concave if for all $x, y, z \in S$ with $x < z < y$,

$$f(z) > \frac{y-z}{y-x} f(x) + \frac{z-x}{y-x} f(y).$$

Therefore, the effective discount rate function is strictly concave if for all $\lambda_0, \lambda_1, \lambda_2 \in$supp$(\Psi)$ with $\lambda_0 < \lambda_2 < \lambda_1$,

$$\widetilde{r}(\lambda_2) > \frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_0}\widetilde{r}(\lambda_0) + \frac{\lambda_2 - \lambda_0}{\lambda_1 - \lambda_0}\widetilde{r}(\lambda_1).$$

Equivalenty,

$$\frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_0} > \frac{\widetilde{r}(\lambda_1) - \widetilde{r}(\lambda_2)}{\widetilde{r}(\lambda_2) - \widetilde{r}(\lambda_0)}.$$

Using (14), and using the fact that the effective discount rate is strictly increasing,

$$\begin{aligned}\frac{\widetilde{r}(\lambda_1)-\widetilde{r}(\lambda_2)}{\widetilde{r}(\lambda_2)-\widetilde{r}(\lambda_0)} &= \frac{\int_0^1 \mu\lambda_1 \frac{\lambda'}{\Lambda}\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda_1)+\widetilde{r}(\lambda')}d\Psi(\lambda') - \int_0^1 \mu\lambda_2 \frac{\lambda'}{\Lambda}\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda_2)+\widetilde{r}(\lambda')}d\Psi(\lambda')}{\int_0^1 \mu\lambda_2 \frac{\lambda'}{\Lambda}\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda_2)+\widetilde{r}(\lambda')}d\Psi(\lambda') - \int_0^1 \mu\lambda_0 \frac{\lambda'}{\Lambda}\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda_0)+\widetilde{r}(\lambda')}d\Psi(\lambda')} \\ &< \frac{\int_0^1 \mu(\lambda_1-\lambda_2)\frac{\lambda'}{\Lambda}\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda_2)+\widetilde{r}(\lambda')}d\Psi(\lambda')}{\int_0^1 \mu(\lambda_2-\lambda_0)\frac{\lambda'}{\Lambda}\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda_2)+\widetilde{r}(\lambda')}d\Psi(\lambda')} = \frac{\lambda_1-\lambda_2}{\lambda_2-\lambda_0}.\end{aligned}$$

Hence, the effective discount rate function is strictly concave.

To derive the last property of the effective discount rate, take the expectation of the equation (14):

$$\begin{aligned}\int_0^1 \widetilde{r}(\lambda)\,d\Psi(\lambda) &= r + \int_0^1\int_0^1 \mu\lambda\frac{\lambda'}{\Lambda}\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')}d\Psi(\lambda')d\Psi(\lambda) \\ &= r + \frac{1}{2}\int_0^1\int_0^1 \mu\lambda\frac{\lambda'}{\Lambda}\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')}d\Psi(\lambda')d\Psi(\lambda) \\ &\quad + \frac{1}{2}\int_0^1\int_0^1 \mu\lambda\frac{\lambda'}{\Lambda}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')}d\Psi(\lambda')d\Psi(\lambda) \\ &= r + \frac{1}{2}\int_0^1\int_0^1 \mu\lambda\frac{\lambda'}{\Lambda}\frac{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')}d\Psi(\lambda')d\Psi(\lambda) \\ &= r + \frac{1}{2}\int_0^1\int_0^1 \mu\lambda\frac{\lambda'}{\Lambda}d\Psi(\lambda')d\Psi(\lambda) \\ &= r + \frac{\mu\Lambda}{2}.\end{aligned}$$

B.4 Proof of Proposition 3

I first take the Fourier transform of the second line of equation (20):

$$\int_{-\infty}^{\infty}\left[\int_{0}^{1}\int_{-1}^{1}\int_{-\infty}^{\infty}2\mu\lambda\frac{\lambda'}{\Lambda}\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)\phi_{\rho,\lambda}(a')\,\phi_{\rho',\lambda'}\left(a\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)-a'+\overline{C}\left[(\rho,\lambda),(\rho',\lambda')\right]\right)\right.$$
$$\left.da'dF(\rho')\,d\Psi(\lambda')\right]e^{-i2\pi az}da$$

$$=\int_{0}^{1}\int_{-1}^{1}\int_{-\infty}^{\infty}2\mu\lambda\frac{\lambda'}{\Lambda}\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)\phi_{\rho,\lambda}(a')$$
$$\left[\int_{-\infty}^{\infty}\phi_{\rho',\lambda'}\left(a\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)-a'+\overline{C}\left[(\rho,\lambda),(\rho',\lambda')\right]\right)e^{-i2\pi az}da\right]da'dF(\rho')\,d\Psi(\lambda')$$

$$=\int_{0}^{1}\int_{-1}^{1}\int_{-\infty}^{\infty}\frac{2\mu\lambda\lambda'}{\Lambda}\phi_{\rho,\lambda}(a')\,e^{\frac{i2\pi z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\left\{-a'+\overline{C}[(\rho,\lambda),(\rho',\lambda')]\right\}}$$
$$\left[\int_{-\infty}^{\infty}\phi_{\rho',\lambda'}\left(a\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)-a'+\overline{C}\left[(\rho,\lambda),(\rho',\lambda')\right]\right)\right.$$
$$\left.e^{\frac{-i2\pi z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\left\{a\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)-a'+\overline{C}[(\rho,\lambda),(\rho',\lambda')]\right\}}d\left(a\left(1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}\right)-a'+\overline{C}\left[(\rho,\lambda),(\rho',\lambda')\right]\right)\right]da'dF(\rho')\,d\Psi(\lambda')$$

$$=\int_{0}^{1}\int_{-1}^{1}\int_{-\infty}^{\infty}2\mu\lambda\frac{\lambda'}{\Lambda}\phi_{\rho,\lambda}(a')\,e^{i2\pi\left\{-a'+\overline{C}[(\rho,\lambda),(\rho',\lambda')]\right\}\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}}\widehat{\phi}_{\rho',\lambda'}\left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\right)da'dF(\rho')\,d\Psi(\lambda')$$

$$= \int_0^1 \int_{-1}^1 2\mu\lambda \frac{\lambda'}{\Lambda} \widehat{\phi}_{\rho',\lambda'} \left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} \right) e^{i2\pi \overline{C}[(\rho,\lambda),(\rho',\lambda')] \frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}}$$

$$\left[\int_{-\infty}^{\infty} \phi_{\rho,\lambda}(a') e^{-i2\pi a' \frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}} da' \right] dF(\rho') d\Psi(\lambda')$$

$$= \int_0^1 \int_{-1}^1 2\mu\lambda \frac{\lambda'}{\Lambda} \widehat{\phi}_{\rho',\lambda'} \left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} \right) e^{i2\pi \overline{C}[(\rho,\lambda),(\rho',\lambda')] \frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}} \widehat{\phi}_{\rho,\lambda} \left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} \right) dF(\rho') d\Psi(\lambda').$$

And using the linearity and integrability of the Fourier transform, equation (24) is obtained.

To obtain equations (25) and (26), I use the identities satisfied by the Fourier transform (see Bracewell, 2000, p. 152-154) for any function $g(x)$

$$\widehat{g}(0) = \int_{-\infty}^{\infty} g(x)dx$$

and

$$\widehat{g}'(0) = -i2\pi \int_{-\infty}^{\infty} xg(x)dx$$

respectively.

n-th conditional moment of asset holdings can be written as follows using the Fourier transform

$$\mathbb{E}_\phi [a^n \mid \rho, \lambda] = (-i2\pi)^{-n} \left[\frac{d^n}{dz^n} \widehat{\phi}_{\rho,\lambda}(z) \right]_{z=0}.$$

Let's first use equation (24) to find an expression for $\frac{d^n}{dz^n}\widehat{\phi}_{\rho,\lambda}(z)$:

$$(\alpha + 2\mu\lambda) \frac{d^n}{dz^n} \widehat{\phi}_{\rho,\lambda}(z) = \alpha \int_{-1}^{1} \frac{d^n}{dz^n} \widehat{\phi}_{\rho',\lambda}(z) dF(\rho')$$

$$+ \int_0^1 \int_{-1}^1 2\mu\lambda \frac{\lambda'}{\Lambda} \frac{d^n}{dz^n} \left\{ e^{i2\pi \overline{C}[(\rho,\lambda),(\rho',\lambda')] \frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}} \widehat{\phi}_{\rho,\lambda} \left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} \right) \widehat{\phi}_{\rho',\lambda'} \left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}} \right) \right\} dF(\rho') d\Psi(\lambda')$$

For the second line, I use the following generalization of the product rule:

$$\frac{d^n}{dx^n}\prod_{i=1}^{3} g_i(x) = \sum_{j_1+j_2+j_3=n} \binom{n}{j_1, j_2, j_3} \prod_{i=1}^{3} \frac{d^{j_i}}{dx^{j_i}} g_i(x),$$

$$(\alpha + 2\mu\lambda) \frac{d^n}{dz^n}\widehat{\phi}_{\rho,\lambda}(z) = \alpha \int_{-1}^{1} \frac{d^n}{dz^n}\widehat{\phi}_{\rho',\lambda}(z)\, dF(\rho') + \int_0^1\int_{-1}^{1} 2\mu\lambda\frac{\lambda'}{\Lambda} \sum_{j_1+j_2+j_3=n} \binom{n}{j_1, j_2, j_3}$$
$$\frac{d^{j_1}}{dz^{j_1}} e^{\overline{C}[(\rho,\lambda),(\rho',\lambda')]\frac{i2\pi z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}} \frac{d^{j_2}}{dz^{j_2}}\widehat{\phi}_{\rho,\lambda}\left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\right) \frac{d^{j_3}}{dz^{j_3}}\widehat{\phi}_{\rho',\lambda'}\left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\right) dF(\rho')\, d\Psi(\lambda'),$$

$$(\alpha + 2\mu\lambda)\,\widehat{\phi}^{(n)}_{\rho,\lambda}(z) = \alpha \int_{-1}^{1} \widehat{\phi}^{(n)}_{\rho',\lambda}(z) dF(\rho') + \int_0^1\int_{-1}^{1} 2\mu\lambda\frac{\lambda'}{\Lambda} \sum_{j_1+j_2+j_3=n} \binom{n}{j_1, j_2, j_3}$$
$$(i2\pi\overline{C}\,[(\rho,\lambda),(\rho',\lambda')])^{j_1} (\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')})^n e^{\overline{C}[(\rho,\lambda),(\rho',\lambda')]\frac{i2\pi z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}}$$
$$\widehat{\phi}^{(j_2)}_{\rho,\lambda}\left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\right) \widehat{\phi}^{(j_3)}_{\rho',\lambda'}\left(\frac{z}{1+\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda)}}\right) dF(\rho')\, d\Psi(\lambda'),$$

$$(\alpha + 2\mu\lambda)\,\widehat{\phi}^{(n)}_{\rho,\lambda}(0) = \alpha \int_{-1}^{1} \widehat{\phi}^{(n)}_{\rho',\lambda}(0) dF(\rho') + \int_0^1\int_{-1}^{1} 2\mu\lambda\frac{\lambda'}{\Lambda} (\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda')})^n$$
$$\sum_{j_1+j_2+j_3=n} \binom{n}{j_1, j_2, j_3} \left\{(i2\pi\overline{C}\,[(\rho,\lambda),(\rho',\lambda')])^{j_1}\widehat{\phi}^{(j_2)}_{\rho,\lambda}(0)\,\widehat{\phi}^{(j_3)}_{\rho',\lambda'}(0)\right\} dF(\rho')\, d\Psi(\lambda').$$

Dividing both sides by $(-i2\pi)^n$:

$$(\alpha+2\mu\lambda)\,\mathbb{E}_\phi\left[a^n \mid \rho,\lambda\right] = \alpha\mathbb{E}_\phi\left[a^n|\lambda\right] + \int_0^1\int_{-1}^1 2\mu\lambda\frac{\lambda'}{\Lambda}\left(\frac{\widetilde{r}\,(\lambda)}{\widetilde{r}\,(\lambda)+\widetilde{r}\,(\lambda')}\right)^n$$

$$\sum_{j_1+j_2+j_3=n}\begin{pmatrix} n \\ j_1,j_2,j_3\end{pmatrix}\left\{\left(-\overline{C}\left[(\rho,\lambda),(\rho',\lambda')\right]\right)^{j_1}\mathbb{E}_\phi\left[a^{j_2}\mid\rho,\lambda\right]\mathbb{E}_\phi\left[a^{j_3}\mid\rho',\lambda'\right]\right\}dF\left(\rho'\right)d\Psi\left(\lambda'\right).$$

Using the multinomial expansion of $\left(-\overline{C}\left[(\rho,\lambda),(\rho',\lambda')\right]\right)^{j_1}$:

$$\begin{aligned}(\alpha+2\mu\lambda)\,\mathbb{E}_\phi\left[a^n \mid \rho,\lambda\right] &= \alpha\mathbb{E}_\phi\left[a^n|\lambda\right] \\ &+ \int_0^1\int_{-1}^1 2\mu\lambda\frac{\lambda'}{\Lambda}\left(\frac{\widetilde{r}\,(\lambda)}{\widetilde{r}\,(\lambda)+\widetilde{r}\,(\lambda')}\right)^n\sum_{j_1+j_2+j_3=n}\begin{pmatrix} n \\ j_1,j_2,j_3\end{pmatrix}\mathbb{E}_\phi\left[a^{j_3}\mid\rho',\lambda'\right] \\ &\mathbb{E}_\phi\left[a^{j_2}\mid\rho,\lambda\right]\sum_{k_1+k_2+k_3=j_1}\begin{pmatrix} j_1 \\ k_1,k_2,k_3\end{pmatrix} \\ &\left(\frac{\sigma_\eta}{\sigma_D}\right)^{k_1+k_2}\left(\frac{-\rho\widetilde{r}\,(\lambda')}{\widetilde{r}\,(\lambda)+\alpha}\right)^{k_1}\left(\frac{\rho'\widetilde{r}\,(\lambda')}{\widetilde{r}\,(\lambda')+\alpha}\right)^{k_2}D\left(\lambda,\lambda'\right)^{k_3}dF\left(\rho'\right)d\Psi\left(\lambda'\right).\end{aligned}$$

$$\begin{aligned}(\alpha+2\mu\lambda)\,\mathbb{E}_\phi\left[a^n \mid \rho,\lambda\right] &= \alpha\mathbb{E}_\phi\left[a^n|\lambda\right] \\ &+ \int_0^1 2\mu\lambda\frac{\lambda'}{\Lambda}\left(\frac{\widetilde{r}\,(\lambda)}{\widetilde{r}\,(\lambda)+\widetilde{r}\,(\lambda')}\right)^n\sum_{j_1+j_2+j_3=n}\begin{pmatrix} n \\ j_1,j_2,j_3\end{pmatrix}\mathbb{E}_\phi\left[a^{j_2}\mid\rho,\lambda\right] \\ &\sum_{k_1+k_2+k_3=j_1}\begin{pmatrix} j_1 \\ k_1,k_2,k_3\end{pmatrix}\left(\frac{\sigma_\eta}{\sigma_D}\right)^{k_1+k_2} \\ &\left(\frac{-\rho\widetilde{r}\,(\lambda')}{\widetilde{r}\,(\lambda)+\alpha}\right)^{k_1}\left(\frac{\widetilde{r}\,(\lambda')}{\widetilde{r}\,(\lambda')+\alpha}\right)^{k_2}D\left(\lambda,\lambda'\right)^{k_3}\mathbb{E}_\phi\left[a^{j_3}\rho^{k_2}\mid\lambda'\right]d\Psi\left(\lambda'\right).\end{aligned}$$

$$
\begin{aligned}
(\alpha + 2\mu\lambda)\, \mathbb{E}_\phi \left[a^n \mid \rho, \lambda\right] &= \alpha \mathbb{E}_\phi \left[a^n | \lambda\right] \\
&+ 2\mu\lambda \sum_{j_1+j_2+j_3=n} \begin{pmatrix} n \\ j_1, j_2, j_3 \end{pmatrix} \mathbb{E}_\phi \left[a^{j_2} \mid \rho, \lambda\right] \sum_{k_1+k_2+k_3=j_1} \begin{pmatrix} j_1 \\ k_1, k_2, k_3 \end{pmatrix} \\
&\left(\frac{-\rho}{\widetilde{r}(\lambda) + \alpha}\right)^{k_1} \left(\frac{\sigma_\eta}{\sigma_D}\right)^{k_1+k_2} \int_0^1 \frac{\lambda'}{\Lambda} \left(\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda) + \widetilde{r}(\lambda')}\right)^n \\
&\widetilde{r}(\lambda')^{k_1} \left(\frac{\widetilde{r}(\lambda')}{\widetilde{r}(\lambda') + \alpha}\right)^{k_2} D(\lambda, \lambda')^{k_3}\, \mathbb{E}_\phi \left[a^{j_3} \rho^{k_2} \mid \lambda'\right] d\Psi(\lambda').
\end{aligned}
$$

Applying the law of iterated expectations, the proof is complete.

B.5 Proof of Lemma 2

Equation (14) implies the system:

$$
\begin{aligned}
\widetilde{r}(\lambda_f) &= r + \mu \frac{\lambda_f \lambda_s}{\Lambda} \frac{\widetilde{r}(\lambda_s)}{\widetilde{r}(\lambda_f) + \widetilde{r}(\lambda_s)} \left(1 - \psi_f\right) + \mu \frac{\lambda_f^2}{2\Lambda} \psi_f\,, \\
\widetilde{r}(\lambda_s) &= r + \mu \frac{\lambda_s^2}{2\Lambda} \left(1 - \psi_f\right) + \mu \frac{\lambda_f \lambda_s}{\Lambda} \frac{\widetilde{r}(\lambda_f)}{\widetilde{r}(\lambda_s) + \widetilde{r}(\lambda_f)} \psi_f\,.
\end{aligned}
$$

Summing up side by side,

$$
\widetilde{r}(\lambda_f) + \widetilde{r}(\lambda_s) = 2r + \mu \frac{\lambda_f^2}{2\Lambda} \psi_f + \mu \frac{\lambda_s^2}{2\Lambda} \left(1 - \psi_f\right) + \mu \frac{\lambda_f \lambda_s}{\Lambda} \frac{\widetilde{r}(\lambda_f)\, \psi_f + \widetilde{r}(\lambda_s) \left(1 - \psi_f\right)}{\widetilde{r}(\lambda_s) + \widetilde{r}(\lambda_f)}.
$$

Using the lemma 1,

$$
\widetilde{r}(\lambda_f) + \widetilde{r}(\lambda_s) = 2r + \mu \frac{\lambda_f^2}{2\Lambda} \psi_f + \mu \frac{\lambda_s^2}{2\Lambda} \left(1 - \psi_f\right) + \mu \frac{\lambda_f \lambda_s}{\Lambda} \frac{r + \frac{\mu\Lambda}{2}}{\widetilde{r}(\lambda_s) + \widetilde{r}(\lambda_f)}.
$$

Then I get the quadratic equation

$$\left(\widetilde{r}\left(\lambda_f\right)+\widetilde{r}\left(\lambda_s\right)\right)^2-\left(2r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{2\Lambda}\right)\left(\widetilde{r}\left(\lambda_f\right)+\widetilde{r}\left(\lambda_s\right)\right)-\mu\frac{\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)=0.$$

Since $\widetilde{r}\left(\lambda_f\right), \widetilde{r}\left(\lambda_s\right)>0$, the relevant solution is

$$\widetilde{r}\left(\lambda_f\right)+\widetilde{r}\left(\lambda_s\right)=r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{4\Lambda}+\sqrt{\left(r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)^2+\mu\frac{\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)}.$$

Combining this with the equation implied by the lemma 1:

$$\psi_f\,\widetilde{r}\left(\lambda_f\right)+\left(1-\psi_f\right)\widetilde{r}\left(\lambda_s\right)=r+\frac{\mu\Lambda}{2},$$

I have a system of two equations in two unknowns. Equivalently, the system can be written as

$$\begin{aligned}
\widetilde{r}\left(\lambda_f\right)\left(1-2\psi_f\right)=&-\left(r+\frac{\mu\Lambda}{2}\right)+\left(1-\psi_f\right)\left(r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)\\
&+\left(1-\psi_f\right)\sqrt{\left(r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)^2+\mu\frac{\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)},\\
\widetilde{r}\left(\lambda_s\right)\left(1-2\psi_f\right)=&\;r+\frac{\mu\Lambda}{2}-\psi_f\left(r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{4\Lambda}+\sqrt{\left(r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)^2+\mu\frac{\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)}\right).
\end{aligned}$$

When $\psi_f \neq \frac{1}{2}$, the system gives the effective discount rates immediately. When $\psi_f = \frac{1}{2}$, I calculate the limit as $\psi_f \to \frac{1}{2}$ using L'Hospital. The resulting effective discount rates are

$$\widetilde{r}(\lambda_f) = \begin{cases} \frac{-\left(r+\frac{\mu\Lambda}{2}\right)+\left(1-\psi_f\right)\left(r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)+\left(1-\psi_f\right)\sqrt{\left(r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)^2+\mu\frac{\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)}}{1-2\psi_f} & \text{if } \psi_f \neq \frac{1}{2} \\ \frac{\lim_{\psi_f\to\frac{1}{2}}\frac{\partial}{\partial\psi_f}\left\{-\left(r+\frac{\mu\Lambda}{2}\right)+\left(1-\psi_f\right)\left(r+\frac{\mu\mathbb{E}\left[\lambda^2\right]}{4\Lambda}+\sqrt{\left(r+\frac{\mu\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)^2+\frac{\mu\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)}\right)\right\}}{-2} & \text{if } \psi_f = \frac{1}{2} \end{cases}$$

and

$$\widetilde{r}(\lambda_s) = \begin{cases} r+\frac{\mu\Lambda}{2}-\psi_f\left(r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)-\psi_f\sqrt{\left(r+\mu\frac{\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)^2+\mu\frac{\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)} & \text{if } \psi_f \neq \frac{1}{2} \\ \frac{\lim_{\psi_f\to\frac{1}{2}}\frac{\partial}{\partial\psi_f}\left\{r+\frac{\mu\Lambda}{2}-\psi_f\left(r+\frac{\mu\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)-\psi_f\sqrt{\left(r+\frac{\mu\mathbb{E}\left[\lambda^2\right]}{4\Lambda}\right)^2+\frac{\mu\lambda_f\lambda_s}{\Lambda}\left(r+\frac{\mu\Lambda}{2}\right)}\right\}}{-2} & \text{if } \psi_f = \frac{1}{2}. \end{cases}$$

Appendix C. Distortion of effective discount rates

To show that for $\lambda' > \lambda$, $\frac{\widetilde{r}^*(\lambda')}{\widetilde{r}(\lambda')} > \frac{\widetilde{r}^*(\lambda)}{\widetilde{r}(\lambda)}$, it suffices to show the effective discount rate function is supermodular in μ and λ. i.e.,

$$\widetilde{r}(\max\{\lambda',\lambda\};\max\{\mu',\mu\})+\widetilde{r}(\min\{\lambda',\lambda\};\min\{\mu',\mu\})\geq\widetilde{r}(\lambda';\mu')+\widetilde{r}(\lambda;\mu\}).$$

Since, in my analysis, μ takes on two values μ and 2μ, the only condition I need to show is

$$\widetilde{r}(\lambda';2\mu)+\widetilde{r}(\lambda;\mu)\geq\widetilde{r}(\lambda;2\mu')+\widetilde{r}(\lambda';\mu).$$

for $\lambda' > \lambda$. In the usual notation

$$\widetilde{r}^*(\lambda')+\widetilde{r}(\lambda)\geq\widetilde{r}^*(\lambda)+\widetilde{r}(\lambda').$$

Equivalently,

$$\int_0^1 2\mu\frac{\lambda''}{\Lambda}\left(\frac{\lambda'\widetilde{r}^*(\lambda'')}{\widetilde{r}^*(\lambda')+\widetilde{r}^*(\lambda'')}-\frac{\lambda\widetilde{r}^*(\lambda'')}{\widetilde{r}^*(\lambda)+\widetilde{r}^*(\lambda'')}\right)d\Psi(\lambda'') - \int_0^1 \mu\frac{\lambda''}{\Lambda}\left(\frac{\lambda'\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda')+\widetilde{r}(\lambda'')}-\frac{\lambda\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda'')}\right)d\Psi(\lambda'')>0.$$

$$\begin{aligned}
&\int_0^1 2\mu\frac{\lambda''}{\Lambda}\left(\frac{\lambda'\widetilde{r}^*(\lambda'')}{\widetilde{r}^*(\lambda')+\widetilde{r}^*(\lambda'')}-\frac{\lambda\widetilde{r}^*(\lambda'')}{\widetilde{r}^*(\lambda)+\widetilde{r}^*(\lambda'')}\right)d\Psi(\lambda'')\\
&-\int_0^1 \mu\frac{\lambda''}{\Lambda}\left(\frac{\lambda'\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda')+\widetilde{r}(\lambda'')}-\frac{\lambda\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda)+\widetilde{r}(\lambda'')}\right)d\Psi(\lambda'')\\
&>\int_0^1 2\mu\frac{\lambda''}{\Lambda}\left(\frac{\lambda'\widetilde{r}^*(\lambda'')}{\widetilde{r}^*(\lambda')+\widetilde{r}^*(\lambda'')}-\frac{\lambda\widetilde{r}^*(\lambda'')}{\widetilde{r}^*(\lambda')+\widetilde{r}^*(\lambda'')}\right)d\Psi(\lambda'')\\
&-\int_0^1 \mu\frac{\lambda''}{\Lambda}\left(\frac{\lambda'\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda')+\widetilde{r}(\lambda'')}-\frac{\lambda\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda')+\widetilde{r}(\lambda'')}\right)d\Psi(\lambda'')\\
&=\int_0^1 2\mu\frac{\lambda''}{\Lambda}(\lambda'-\lambda)\frac{\widetilde{r}^*(\lambda'')}{\widetilde{r}^*(\lambda')+\widetilde{r}^*(\lambda'')}d\Psi(\lambda'')-\int_0^1 \mu\frac{\lambda''}{\Lambda}(\lambda'-\lambda)\frac{\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda')+\widetilde{r}(\lambda'')}d\Psi(\lambda'')\\
&=\int_0^1 \mu\frac{\lambda''}{\Lambda}(\lambda'-\lambda)\left(\frac{2\widetilde{r}^*(\lambda'')}{\widetilde{r}^*(\lambda')+\widetilde{r}^*(\lambda'')}-\frac{\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda')+\widetilde{r}(\lambda'')}\right)d\Psi(\lambda'')\\
&=\frac{\lambda'-\lambda}{\lambda'}\int_0^1 \mu\frac{\lambda''}{\Lambda}\lambda'\left(\frac{2\widetilde{r}^*(\lambda'')}{\widetilde{r}^*(\lambda')+\widetilde{r}^*(\lambda'')}-\frac{\widetilde{r}(\lambda'')}{\widetilde{r}(\lambda')+\widetilde{r}(\lambda'')}\right)d\Psi(\lambda'')\\
&=\frac{\lambda'-\lambda}{\lambda'}\left(\widetilde{r}^*(\lambda')-\widetilde{r}(\lambda')\right)>0.
\end{aligned}$$

Appendix D. Calculation of targeted moments

Consider an investor with M idiosyncratic correlation shocks between time T_0 and T, e.g., at times $T^{(M)} = (T_1, T_2, ..., T_M)$, with $0 \leq T_0 < T_1 < T_2 < ... < T_M < T$. Suppose that the correlation type of this investor is ρ_m during $[T_m, T_{m+1})$, and her search efficiency is λ. Then, the investor's target asset holding for the period $[T_m, T_{m+1})$ is

$$a_{tar}^{(m)} = A - \frac{\sigma_\eta}{\sigma_D} \frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda) + \alpha} (\rho_m - \overline{\rho}).$$

And, suppose that this investor receives N_m trading opportunities between time T_m and time T_{m+1}, e.g., at times $t^{(m)} = \left(t_1^{(m)}, t_2^{(m)}, ..., t_{N_m}^{(m)}\right)$, with $T_m \leq t_1^{(m)} < t_2^{(m)} < ... < t_{N_m}^{(m)} < T_{m+1}$. Given the initial holding a_{0_m} and a realization $t^{(m)} \in [T_m, T_{m+1})^{N_m}$, the time-path of the investor's asset holding during $[T_m, T_{m+1})$ is described by a function $k_m : [T_m, T_{m+1}) \to \mathbb{R}$ defined by

$$\mathbf{k}_m(x) = \begin{cases} a_0^{(m)} & \text{for } T_m \leq x < t_1^{(m)} \\ a_1^{(m)} & \text{for } t_1^{(m)} \leq x < t_2^{(m)} \\ ... & ... \\ a_{N_m}^{(m)} & \text{for } t_{N_m}^{(m)} \leq x < T_{m+1}, \end{cases}$$

where $a_n^{(m)}$ is the post-trade asset holding at time $t_n^{(m)}$ for $n = 1, ..., N_m$. Given the initial holding $a_0^{(m)}$ at T_m, the realized path for an investor's holding during $[T_m, T_{m+1})$ is completely described by the number of contacts, N_m, the vector of contact times, $t^{(m)} \in [T_m, T_{m+1})^{N_m}$, and the vector of post-trade holdings at those contact times, $a^{(m)} = \left(a_1^{(m)}, a_2^{(m)}, ..., a_{N_m}^{(m)}\right) \in \mathbb{R}^{N_m}$. Given $a_0^{(m)}$ and $a^{(m)}$, define the investor's accumulated volume of purchases during $[T_m, T_{m+1})$,

$$\mathbb{V}^p\left(a_0^{(m)}, \mathbf{a}^{(m)}\right) = \sum_{n=1}^{N_m} \max\left\{a_n^{(m)} - a_{n-1}^{(m)}, 0\right\},$$

the accumulated volume of sales,

$$\mathbb{V}^s\left(a_0^{(m)},\mathbf{a}^{(m)}\right)=-\sum_{n=1}^{N_m}\min\left\{a_n^{(m)}-a_{n-1}^{(m)},0\right\},$$

and the (signed) net trade,

$$\begin{aligned}\mathbb{V}^n\left(a_0^{(m)},\mathbf{a}^{(m)}\right) &= \mathbb{I}_{\left\{a_{tar}^{(m)}\geq a_0^{(m)}\right\}}\min\left\{\mathbb{V}^p\left(a_0^{(m)},\mathbf{a}^{(m)}\right),a_{tar}^{(m)}-a_0^{(m)}\right\}\\ &\quad -\mathbb{I}_{\left\{a_{tar}^{(m)}<a_0^{(m)}\right\}}\min\left\{\mathbb{V}^s\left(a_0^{(m)},\mathbf{a}^{(m)}\right),a_0^{(m)}-a_{tar}^{(m)}\right\}.\end{aligned}$$

Then,

$$\mathbb{X}\left(a_0^{(m)},\mathbf{a}^{(m)}\right)=\mathbb{V}^p\left(a_0^{(m)},\mathbf{a}^{(m)}\right)+\mathbb{V}^s\left(a_0^{(m)},\mathbf{a}^{(m)}\right)-\left|\mathbb{V}^n\left(a_0^{(m)},\mathbf{a}^{(m)}\right)\right|$$

measures the volume of assets that are purchased and sold by the investor for intermediation purposes during the time interval $[T_m,T_{m+1})$. In reality, an econometrician, who observes transaction-level data, would not be able to calculate the net trade as she could not observe the target holding of the investor. Alternatively, she would match round-trip trades, which in turn yield a proxy for $X\left(a_0^{(m)},\mathbf{a}^{(m)}\right)$, as Green et al. (2007) and Li and Schürhoff (2012) did. Since I observe the target position of investors implied by my model, I am using the target position to calculate the intermediation volume perfectly.

Similarly, I define, for the period $[T_m,T_{m+1})$, the total trading profit

$$\mathbf{\Pi}^{(m)}=\sum_{n=1}^{N_m}\left(a_n^{(m)}-a_{n-1}^{(m)}\right)P_n,$$

average purchase price

$$\mathbb{P}_p^{(m)}=\left(\sum_{n=1}^{N_m}\max\left\{a_n^{(m)}-a_{n-1}^{(m)},0\right\}P_n\right)/\mathbb{V}^p\left(a_0^{(m)},\mathbf{a}^{(m)}\right),$$

average sale price

$$\mathbb{P}_s^{(m)} = \left(-\sum_{n=1}^{N_m} \min \left\{ a_n^{(m)} - a_{n-1}^{(m)}, 0 \right\} P_n \right) / \mathbb{V}^s \left(a_0^{(m)}, \mathbf{a}^{(m)} \right),$$

and intermediation profit

$$\mathbf{\Pi}_X^{(m)} = \mathbf{\Pi}^{(m)} + \mathbb{I}_{\left\{ \mathbb{V}^n \left(a_0^{(m)}, \mathbf{a}^{(m)} \right) \geq 0 \right\}} \left| \mathbb{V}^n \left(a_0^{(m)}, \mathbf{a}^{(m)} \right) \right| \mathbb{P}_p^{(m)} - \mathbb{I}_{\left\{ \mathbb{V}^n \left(a_0^{(m)}, \mathbf{a}^{(m)} \right) < 0 \right\}} \left| \mathbb{V}^n \left(a_0^{(m)}, \mathbf{a}^{(m)} \right) \right| \mathbb{P}_s^{(m)}.$$

Based on the definitions above, the intermediation markup, defined as a fraction of the mean of the equilibrium price distribution, for the period $[T_0, T)$ is

$$markup_{[T_0,T)} = \left(\sum_{m=1}^{M} \mathbf{\Pi}_X^{(m)} \right) / \left(\sum_{m=1}^{M} \mathbb{X} \left(a_0^{(m)}, \mathbf{a}^{(m)} \right) / 2 \right) / \mathbb{E}_\phi [P].$$

Appendix E. Individual welfare creation

$$\mathbb{W}(\lambda) = \int_0^\infty e^{-rt} \left\{ \int_{-\infty}^{\infty} \int_{-1}^{1} u(\rho, a) \Phi_{\lambda,t}(d\rho, da) \right\} dt = \frac{1}{r} \int_{-\infty}^{\infty} \int_{-1}^{1} u(\rho, a) \Phi_\lambda(d\rho, da).$$

The definition of $u(\rho, a)$ implies that

$$\mathbb{W}(\lambda) = \frac{m_D}{r} A - \frac{\gamma \sigma_D^2}{2} \left(A^2 + var_\phi [a|\lambda] \right) - \gamma \sigma_D \sigma_\eta \left(\overline{\rho} A + cov_\phi [\rho, a|\lambda] \right).$$

From Proposition 3,

$$\begin{aligned}\mathbb{W}(\lambda) = {} & \frac{m_D}{r}A - \frac{\gamma\sigma_D^2}{2}A^2 - \gamma\sigma_D\sigma_\eta\overline{\rho}A \\ & - \frac{\gamma\sigma_D^2}{2}\left(\frac{r}{\widetilde{r}(\lambda)}var_\phi[a|\lambda] + \frac{\widetilde{r}(\lambda)-r}{\widetilde{r}(\lambda)}\left(\frac{\sigma_\eta}{\sigma_D}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\right)^2 var[\rho]\right) \\ & + \gamma\sigma_\eta^2 var[\rho]\frac{2(\widetilde{r}(\lambda)-r)}{\alpha+2(\widetilde{r}(\lambda)-r)}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}.\end{aligned}$$

By rearranging,

$$\begin{aligned}\mathbb{W}(\lambda) = {} & \frac{m_D}{r}A - \frac{\gamma\sigma_D^2}{2}A^2 - \gamma\sigma_D\sigma_\eta\overline{\rho}A \\ & + \gamma\sigma_\eta^2 var[\rho]\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\left(\frac{2(\widetilde{r}(\lambda)-r)}{\alpha+2(\widetilde{r}(\lambda)-r)} - \frac{1}{2}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\right) \\ & - \frac{\gamma\sigma_D^2}{2}\frac{r}{\widetilde{r}(\lambda)}\left(var_\phi[a|\lambda] - \left(\frac{\sigma_\eta}{\sigma_D}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\right)^2 var[\rho]\right).\end{aligned}$$

Using the definition of θ,

$$\begin{aligned}\mathbb{W}(\lambda) = {} & \frac{m_D}{r}A - \frac{\gamma\sigma_D^2}{2}A^2 - \gamma\sigma_D\sigma_\eta\overline{\rho}A \\ & + \gamma\sigma_\eta^2 var[\rho]\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\left(\frac{2(\widetilde{r}(\lambda)-r)}{\alpha+2(\widetilde{r}(\lambda)-r)} - \frac{1}{2}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\right) \\ & - \frac{\gamma\sigma_D^2}{2}\frac{r}{\widetilde{r}(\lambda)}\left(var_\phi[\theta|\lambda] - \frac{2\alpha}{\alpha+2(\widetilde{r}(\lambda)-r)}\left(\frac{\sigma_\eta}{\sigma_D}\frac{\widetilde{r}(\lambda)}{\widetilde{r}(\lambda)+\alpha}\right)^2 var[\rho]\right).\end{aligned}$$

References

[1] Afonso, G. (2011). Liquidity and congestion. Journal of Financial Intermediation, 20(3), 324–360.

[2] Afonso, G., Kovner, A., & Schoar, A. (2013). Trading partners in the interbank lending market. Federal Reserve Bank of New York Staff Report.

[3] Afonso, G., & Lagos, R. (2012). An empirical study of trade dynamics in the fed funds market. Federal Reserve Bank of New York Staff Report.

[4] Afonso, G., & Lagos, R. (2015). Trade dynamics in the market for federal funds. Econometrica, 83, 263-313.

[5] Andrei, D., & Cujean, J. (2014). Information percolation, momentum and reversal. Mimeo.

[6] Ashcraft, A., & Duffie, D. (2007). Systemic illiquidity in the federal funds market. American Economic Review, Papers and Proceedings, 97, 221–225.

[7] Atkeson, A. G., Eisfeldt, A. L. & Weill, P.-O. (2015). Entry and exit in OTC derivatives markets. Econometrica, Forthcoming.

[8] Babus, A., & Kondor, P. (2012). Trading and information diffusion in OTC markets. Mimeo.

[9] Bech, M., & Atalay, E. (2010). The topology of the federal funds market. Physica A, 389, 5223–5246.

[10] Biais, B. (1993). Price formation and equilibrium liquidity in fragmented and centralized markets. Journal of Finance, 48, 157–185.

[11] Bracewell, R. N. (2000). The Fourier transform and its applications. New York, NY: McGraw Hill.

[12] Chang, B., & Zhang, S. (2015). Endogenous market making and network formation. Mimeo.

[13] Colliard, J.-E., & Demange, G. (2014). Cash providers: Asset dissemination over intermediation chains. Mimeo.

[14] Constantinides, G. M. (1986). Capital market equilibrium with transaction costs. Journal of Political Economy, 94, 842–862.

[15] Cujean, J. & Praz, R. (2015). Asymmetric information and inventory concerns in over-the-counter markets. Mimeo.

[16] Duffie, D. (2012a). Dark markets: Asset pricing and information transmission in over-the-counter markets. Princeton, NJ: Princeton University Press.

[17] Duffie, D. (2012b). Market making under the proposed Volcker rule. Mimeo.

[18] Duffie, D., Gârleanu, N., & Pedersen, L. H. (2005). Over-the-counter markets. Econometrica, 73, 1815–1847.

[19] Duffie, D., Gârleanu, N., & Pedersen, L. H. (2007). Valuation in over-the-counter markets. Review of Financial Studies, 20, 1865–1900.

[20] Duffie, D., Giroux, G., & Manso, G. (2010). Information percolation. American Economic Journal: Microeconomics, 2, 100-111.

[21] Duffie, D., Malamud, S., & Manso, M. (2009). Information percolation with equilibrium search dynamics. Econometrica, 77(5), 1513–1574.

[22] Duffie, D., & Manso, M. (2007). Information percolation in large markets. American Economic Review, Papers and Proceedings, 97, 203-209.

[23] Farboodi, M. (2014). Intermediation and voluntary exposure to counterparty risk. Mimeo.

[24] Farboodi, M., Jarosch, G., & Shimer, R. (2015). Meeting technologies in decentralized asset markets. Mimeo.

[25] Friewald, N., Jankowitsch, R., & Subrahmanyam, M. G. (2012). Illiquidity or credit deterioration: A study of liquidity in the US corporate bond market during financial crises. Journal of Financial Economics, 105(1), 18-36.

[26] Gârleanu, N. (2009). Portfolio choice and pricing in illiquid markets. Journal of Economic Theory, 144(2), 532–564.

[27] Gavazza, A. (2011a). Leasing and secondary markets: Theory and evidence from commercial aircraft. Journal of Political Economy, 119(2), 325–377.

[28] Gavazza, A. (2011b). The role of trading frictions in real asset markets. American Economic Review, 101(4), 1106–1143.

[29] Gofman, M. (2011). A network-based analysis of over-the-counter markets. Mimeo.

[30] Green, R. C., Hollifield, B., & Schürhoff, N. (2007). Financial intermediation and the costs of trading in an opaque market. Review of Financial Studies, 20, 275-314.

[31] Hollifield, B., Neklyudov, A., & Spatt, C. S. (2014). Bid-ask spreads and the pricing of securitizations:144a vs. registered securitizations. Mimeo.

[32] Hugonnier, J., Lester, B. & Weill, P.-O. (2014). Heterogeneity in decentralized asset markets. Mimeo.

[33] Krasnosel'skiĭ, M. A. (1964). Positive solutions of operator equations. Groningen, the Netherlands: P. Noordhoff Ltd.

[34] Lagos, R., & Rocheteau, G. (2007). Search in asset markets: Market structure, liquidity, and welfare. American Economic Review, Papers and Proceedings, 97, 198–202.

[35] Lagos, R., & Rocheteau, G. (2009). Liquidity in asset markets with search frictions. Econometrica, 77, 403–426.

[36] Lagos, R., Rocheteau, G., & Weill, P.-O. (2011). Crises and liquidity in over-the-counter markets. Journal of Economic Theory, 146(6), 2169–2205.

[37] Lester, B., Rocheteau, G. & Weill, P.-O. (2015). Competing for order flow in OTC markets. Journal of Money, Credit and Banking, 47, 77-126.

[38] Li, D., & Schürhoff, N. (2012). Dealer networks. Mimeo.

[39] Malamud, S., & Rostek, M. (2012). Decentralized exchange. Mimeo.

[40] Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). Microeconomic theory. Oxford, UK: Oxford University Press.

[41] Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory, 3(4), 373-413.

[42] Municipal Securities Rulemaking Board. (2008). Fact Book.

[43] Neklyudov, A. (2015). Bid-ask spreads and the over-the-counter interdealer markets: Core and peripheral dealers. Mimeo.

[44] Pagnotta, S. E., & Philippon, T. (2015). Competing on speed. Mimeo.

[45] Praz, R. (2014). Equilibrium asset pricing with both liquid and illiquid markets. Mimeo.

[46] Protter, P. (2004). Stochastic integration and differential equations. New York, NY: Springer.

[47] Randall, O. (2015). Pricing and liquidity in over-the-counter markets. Mimeo.

[48] Shen, J., Wei, B., & Yan, H. (2015). Financial intermediation chains in an OTC market. Mimeo.

[49] Shimer, R., & Smith, L. (2001). Matching, search, and heterogeneity. The B.E. Journal of Macroeconomics, 1(1), 1-18.

[50] Siriwardane, E. N. (2015). Concentrated capital losses and the pricing of corporate credit risk. Mimeo.

[51] Vayanos, D., & Wang, T. (2007). Search and endogenous concentration of liquidity in asset markets. Journal of Economic Theory, 136, 66-104.

[52] Vayanos, D., & Weill, P.-O. (2008). A search-based theory of the on-the-run phenomenon. Journal of Finance, 63, 1361–1398.

[53] Weill, P.-O. (2007). Leaning against the wind. Review of Economic Studies, 74(4), 1329–1354.

[54] Weill, P.-O. (2008). Liquidity premia in dynamic bargaining markets. Journal of Economic Theory, 140, 66–96.

[55] Yüceer, Ü. (2002). Discrete convexity: convexity for functions defined on discrete spaces. Discrete Applied Mathematics, 119, 297-304.

CHAPTER 2

Price Dispersion and Trading Activity during Turbulent Times

1 Introduction

I propose a search-and-bargaining model of crises in decentralized asset markets *à la* Duffie, Gârleanu and Pedersen (2005) where risk-averse investors with time-varying hedging needs randomly contact each other and bargain bilaterally over the terms of trade including price and quantity. The crisis is modeled as a one-time aggregate shock to uncertainty with a random recovery. The arrival of the crisis shock leads to an increase in both the volatility of asset payoff and the volatility of investors' background risk.

The equilibrium path for investors' valuations, terms of trade, and the distribution of investors' positions is derived in closed form both during the crisis and during the recovery. Having unrestricted asset positions in the model appears quite difficult since it leads to a new endogenous dimension of the equilibrium objects. As trade happens bilaterally during random contacts, there exists a non-trivial heterogeneity in investors' asset holdings, even among the ones who are identical with respect to hedging needs. Keeping track of this heterogeneity complicates the equilibrium computation. However, using the convolution methods as in Chapter 1, I show that the model is fully tractable. Tractability of the model allows me to derive natural proxies for price dispersion and trading activity.

The existence of search frictions leads to distortion in holdings on the extensive margin, which in turn creates dispersion in marginal valuations. When each pair of buyer and seller contact, their negotiated price depends on their current marginal valuations. This gives rise to equilibrium price dispersion. I show that both volatility of asset payoff and volatility of background risk contribute to higher level of price dispersion during the crisis. This is consistent with the empirical evidence presented by Friewald, Jankowitsch and Subrahmanyam (2012) for the US corporate bond market and Afonso and Lagos (2012) for the federal funds market that the price dispersion is higher during crises compared to normal market periods.

The effect of crisis on trading activity might be different in the long term and in the short term, depending on the relative positioning of the change in the volatility of asset payoff and in the volatility of background risk. When the increase in the volatility of asset payoff is higher than a certain threshold, in the long term, a flight from this market is observed. This is consistent with the "flight-to-quality" events observed in real-life financial markets during crises. When overall uncertainty in the markets is high, investors trade mostly in relatively safe markets. Similarly, when the volatility of asset payoff is lower than the threshold, there is flight-to-quality to this market. Regarding the short term, I find that a flight to the asset market always starts with a “heating-up” in trading activity but a flight from the market might start with a dry-up or heating-up during the onset of the crisis. If the relative increase in the volatility of asset payoff is too high, a period of fire sales is triggered leading to a short heating-up before the complete dry-up of the trading activity. Therefore, very severe flights from the market actually starts with a heating-up where fire sales occur in which investors quickly transition to more cautious positions in the asset.

I calibrate the model according to the US corporate bond market data and show the effect of this aggregate uncertainty shock created by the subprime crisis on price dispersion and trading activity. Since the calibrated values imply that the volatility of corporate bond payoffs increased more than the increase in the overall volatility of markets, we observe a flight-to-quality from the corporate bond market during the aggregate uncertainty shock in the long term. On the other hand, as the difference between the increase in the volatility of corporate bond payoffs and the increase in the overall volatility of markets is sufficiently high, a heating-up during the onset of the crisis is observed. Because the investors’ new trading regime after the aggregate shock is much more cautious than their behavior during the normal times, they have a strong incliniation to holding conservative positions after the aggregate uncertainty shock. This leads to plenty of fire sales initially. This, in turn, creates a spike in price dispersion and a heating-up in the trading activity. The short heating-up period is followed by a long dry-up in the trading activity, which lasts until the recovery.

This paper belongs to search-theoretic asset pricing literature spurred by Duffie et al. (2005). More precisely, it studies an aggregate shock similar to Duffie, Gârleanu and Pedersen (2007), Weill (2008), and Lagos, Rocheteau and Weill (2011). Investors in the model of Duffie et al. (2007) have binary valuation for holding an indivisible asset. Relative to their model, my model features an arbitrary heterogeneity in valuations and unrestricted divisible holdings of the asset. Heterogeneity in valuations creates price dispersion in equilibrium, while investors' optimal holding decisionof the divisible asset provide them with flexibility to respond to changes in market conditions. Both of these additional features are essential to my analysis of the effect of the crisis on price dispersion and trading activity. Weill (2008) and Lagos et al. (2011) study a partially centralized market in which investors are able to trade infrequently by paying an intermediation fee to exogenously designated dealers. The difference of my model is that it features endogenous intermediation. Therefore, changes in trading activity captures the changes in trades for intermediation purposes as well. This allows an analysis of intermediation chains.

The remainder of the paper is organized as follows. Section 2 describes the model environment. Section 3 studies the equilibrium of this environment, while Section 4 assesses the main results. Section 5 concludes.

2 Environment

Time is continuous and runs forever. 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 (see Protter, 2004). There is a continuum of investors with a total measure normalized to 1. There is one long-lived asset in fixed supply denoted by A. This asset is traded over the counter, and pays an expected dividend flow denoted by m_D. There is also a perishable good, called the *numéraire*, which all investors produce and consume.

The specification of preferences and trading motives are exactly same as Chapter 1. Investors' level of risk aversion and time preference rate are denoted by γ and r respectively. The instantaneous utility function of an investor is $u(\rho, a, t) + c$, where

$$u(\rho, a, t) \equiv a m_D - \frac{1}{2} r\gamma \left(a^2 \sigma_D^2(t) + 2\rho a \sigma_D(t) \sigma_\eta(t) \right) \tag{1}$$

is the instantaneous quadratic benefit to the investor from holding $a \in \mathbb{R}$ units of the asset when of type $\rho \in [-1, +1]$, and $c \in \mathbb{R}$ denotes the net consumption of the numéraire good. An investor's net consumption becomes negative when she produces the numéraire to make side payments.

This utility specification is interpreted in terms of risk aversion. Since the parameter m_D is an expected rather than actual dividend flow, this cash flow needs to be adjusted for risk. The term $a^2\sigma_D^2(t)$ represents the instantaneous variance of the asset payoff where $\sigma_D(t)$ is the volatility of the asset payoff. The term $2\rho a \sigma_D(t) \sigma_\eta(t)$ captures the instantaneous covariance between the asset payoff and some background risk with volatility $\sigma_\eta(t)$. Therefore, the investor's type ρ captures the instantaneous correlation between the asset payoff and the background risk. Duffie et al. (2007), Vayanos and Weill (2008), Gârleanu (2009), and Chapter 1 of this dissertation show that this quadratic utility specification can be derived from first principles.[1] I keep the microfoundation of this specification out of the paper because the reduced-form imparts the main intuitions without the burden of derivations.

For brevity of the notation later, I define the parameters

$$\begin{aligned}\kappa_0 &\equiv m_D, \\ \kappa_1(t) &\equiv r\gamma\sigma_D^2(t), \\ \kappa_2(t) &\equiv r\gamma\sigma_D(t)\,\sigma_\eta(t).\end{aligned}$$

The correlation between the asset payoff and the background risk is heterogeneous across investors. Investors occasionally receive idiosyncratic shocks to their correlation type, which creates the motive to trade. Arrival of these shocks is independent from other stochastic processes and across investors. The idiosyncratic shocks occur at random Poisson arrival times with intensity $\alpha > 0$, and upon the arrival of an idiosyncratic shock, the investor's new type is drawn according to the cdf F on $[-1, +1]$. Time-dependence of $\kappa_1(t)$ and $\kappa_2(t)$ captures a common shock to the level of uncertainty in the OTC market and the economy. This is designed to capture a crisis period in the spirit of Lagos et al. (2011). More specifically, I assume that the economy is initially in its stationary equilibrium, where $\kappa_1(t) = \kappa_1 \equiv r\gamma\sigma_D^2$ and $\kappa_2(t) = \kappa_2 \equiv r\gamma\sigma_D\sigma_\eta$. At date 0, an unexpected aggregate shock hits the economy, after which $\kappa_1(t) = \overline{\kappa}_1 \equiv r\gamma\overline{\sigma}_D^2 > \kappa_1$ and $\kappa_2(t) = \overline{\kappa}_2 \equiv r\gamma\overline{\sigma}_D\overline{\sigma}_\eta > \kappa_2$. Investors expect the recovery to arrive as a one-time Poisson event with intensity $R > 0$.

Under the new brief notation, investors' utility flow during the stationary equilibrium and after the recovery is

$$u(\rho, a) \equiv \kappa_0 a - \frac{1}{2}\kappa_1 a^2 - \kappa_2 \rho a$$

[1] Appendix A of Chapter 1 assumes that investors have CARA preferences over the numéraire good, and they can invest in a riskless asset traded in a Walrasian market, and in a risky asset traded over the counter. Moreover, the investor receives a random income whose correlation with the payoff of risky asset is ρ. These assumptions give rise to the reduced-form specification, up to a suitable first-order approximation.

and during the crisis period it is

$$\overline{u}(\rho, a) \equiv \kappa_0 a - \frac{1}{2}\overline{\kappa}_1 a^2 - \overline{\kappa}_2 \rho a,$$

where $\overline{\kappa}_1/\kappa_1$ captures the severity of the crisis in the OTC market for the traded asset and $\overline{\kappa}_2/\kappa_2$ captures the severity of the crisis in the combination of all other markets.

All trades are fully bilateral. Pair-wise meetings among investors follow standard random search and matching dynamics. A given investors meets another investor at random Poisson arrival times with intensity $\lambda > 0$, reflecting the overall search efficiency of the market. Conditional on a meeting, the counterparty is drawn randomly from the pool of all investors. A meeting between investor (ρ, a) and investor (ρ', a') at date t is followed by a bargaining process over quantity q and unit price P. The specific bargaining protocol I employ is the axiomatic bargaining *à la* Nash (1950) in which investors are symmetric in their bargaining strengths. The resulting number of assets that the investor (ρ, a) purchases is denoted by $q\left[(\rho, a), (\rho', a'), t\right]$. Thus, she will become an investor of type $(\rho, a + q\left[(\rho, a), (\rho', a'), t\right])$ after this trade, while her counterparty will become type $(\rho', a' - q\left[(\rho, a), (\rho', a'), t\right])$. The per unit price, the investor (ρ, a) will pay, is denoted by $P\left[(\rho, a), (\rho', a'), t\right]$.

3 Equilibrium

In this section, I solve for the equilibrium of this economy in two steps. First, I characterize the equilibrium path after the recovery has occurred. To do so, I take as given the recovery time T_R and the joint distribution $\Phi\left(\rho, a | T_R\right)$ of types and asset holdings at $t = T_R$. Then, I characterize the equilibrium path during the crisis, i.e. when $t \in [0, T_R)$.

3.1 Equilibrium path after the recovery

3.1.1 Investor's problem

Let me start by describing the investor's problem after the recovery shock has occurred at $t = T_R$. Suppose $J(\rho, a, t)$ denotes the maximum attainable continuation utility of an investor with type ρ and asset holding a at date t. Taking the pricing function $P\left[(\rho, a), (\rho', a'), t\right]$ and the trade size function $q\left[(\rho, a), (\rho', a'), t\right]$ as given, $J(\rho, a, t)$ is defined as

$$J(\rho, a, t) \equiv \mathbb{E}_t \left[\int_t^{\infty} e^{-r(s-t)} \left[u(\rho_s, a_s) + c_s \right] ds \mid \rho_t = \rho,\ a_t = a \right],$$

s.t.

$$da_t = \begin{cases} q\left[(\rho_t, a_t), (\rho'_t, a'_t), t\right] & \text{if there is contact with investor } (\rho'_t, a'_t) \\ 0 & \text{if no contact,} \end{cases}$$

$$c_t = \begin{cases} -P\left[(\rho_t, a_t), (\rho'_t, a'_t), t\right] da_t & \text{if there is contact with investor } (\rho'_t, a'_t) \\ 0 & \text{if no contact.} \end{cases}$$

Taking the derivative with respect to t and rearranging, one can show that $J(\rho, a, t)$ satisfies the following ordinary differential equation (ODE):

$$\begin{aligned} J_t(\rho, a, t) = {} & rJ(\rho, a, t) - u(\rho, a) - \alpha \int_{-1}^{1} [J(\rho', a, t) - J(\rho, a, t)] dF(\rho') \\ & - 2\lambda \int_{-\infty}^{\infty} \int_{-1}^{1} \{ J(\rho, a + q\left[(\rho, a), (\rho', a'), t\right], t) - J(\rho, a, t) \\ & - q\left[(\rho, a), (\rho', a'), t\right] P\left[(\rho, a), (\rho', a'), t\right] \} \Phi(d\rho', da' | t). \end{aligned} \quad (2)$$

The first term on the right hand side of the equation (2) is the investor's "flow" continuation utility. The difference between this flow continuation utility and the instantaneous change in her current status governs the growth of the continuation utility. The second term captures the utility flow from her current holding. The third term is the expected change in the continuation utility, conditional on switching preference types, which occurs with intensity α; and the fourth term is the expected change in the continuation utility, conditional on trade, which occurs with intensity 2λ, since an investor finds others at rate λ and are found by others at rate λ. Conditional on contact, the counterparty is drawn randomly from the distribution of types and asset holdings with cdf $\Phi(\rho', a'|t)$. Terms of trade, $q\left[(\rho, a), (\rho', a'), t\right]$ and $P\left[(\rho, a), (\rho', a'), t\right]$, maximize the symmetric Nash product (3), subject to usual individual rationality constraints.

$$\begin{aligned} &\left[q\left[(\rho, a), (\rho', a'), t\right], P\left[(\rho, a), (\rho', a'), t\right]\right] \\ &= \arg\max_{q,P} \left[J(\rho, a+q, t) - J(\rho, a, t) - Pq\right]^{\frac{1}{2}} \left[J(\rho', a'-q, t) - J(\rho', a', t) + Pq\right]^{\frac{1}{2}}, \end{aligned} \quad (3)$$

s.t.

$$J(\rho, a+q, t) - J(\rho, a, t) - Pq \geq 0,$$
$$J(\rho', a'-q, t) - J(\rho', a', t) + Pq \geq 0.$$

3.1.2 Trades

Before solving for the path for the equilibrium distribution of asset holdings, I will characterize the trades happening during the recovery path. To do so, I start by solving the optimization problem (3) of Nash bargaining. Solving this problem is relatively straightforward: I set up the Lagrangian of this problem. Then I find the first-order necessary and sufficient conditions (see Theorem M.K.2., p. 959, and Theorem M.K.3., p. 961, in Mas-

Colell, Whinston & Green, 1995) for optimality by differentiating the Lagrangian. The trade size $q\left[(\rho,a),(\rho',a'),t\right]$ solves

$$J_a(\rho,a+q,t)=J_a(\rho',a'-q,t). \tag{4}$$

Notice that the quantity which solves the equation (4) is also the maximizer of the total trade surplus, i.e.

$$q\left[(\rho,a),(\rho',a'),t\right]=\arg\max_q\ J(\rho,a+q,t)-J(\rho,a,t)+J(\rho',a'-q,t)-J(\rho',a',t).$$

Continuous differentiability and strict concavity of $J(\rho,.,t)$ for all ρ and t, which I will establish later, guarantees the existence and uniqueness of $q\left[(\rho,a),(\rho',a'),t\right]$. Then, the transaction price $P\left[(\rho,a),(\rho',a'),t\right]$ is determined such that the total trade surplus is split equally between the parties

$$P=\frac{J(\rho,a+q,t)-J(\rho,a,t)-(J(\rho',a'-q,t)-J(\rho',a'),t)}{2q} \tag{5}$$

if $J_2(\rho,a,t)\neq J_2(\rho',a',t)$; and $P=J_2(\rho,a,t)$ if $J_2(\rho,a,t)=J_2(\rho',a',t)$. Substituting the trade quantity and price into (2), I get

$$J_t(\rho,a,t)=rJ(\rho,a,t)-u(\rho,a)-\alpha\int_{-1}^{1}[J(\rho',a,t)-J(\rho,a,t)]dF(\rho')$$
$$-\lambda\int_{-\infty}^{\infty}\int_{-1}^{1}\left[\max_q\ \{J(\rho,a+q,t)-J(\rho,a,t)+J(\rho',a'-q,t)-J(\rho',a',t)\}\right]\Phi(d\rho',da'|t). \tag{6}$$

To solve for $J(\rho,a,t)$, I follow a guess&verify approach as in the first chapter. My conjecture is

$$J(\rho,a,t)=D\left(t\right)+E\left(t\right)\rho+F\left(t\right)a+G\left(t\right)a^{2}+H\left(t\right)\rho a+M\left(t\right)\rho^{2}.$$

Applying (4),

$$F\left(t\right)+2G\left(t\right)\left(a+q\right)+H\left(t\right)\rho=F\left(t\right)+2G\left(t\right)\left(a'-q\right)+H\left(t\right)\rho'.$$

Then,

$$q\left[\left(\rho,a\right),\left(\rho',a'\right),t\right]=\frac{a'-a}{2}+\frac{H\left(t\right)\left(\rho'-\rho\right)}{4G\left(t\right)}$$

and

$$P\left[\left(\rho,a\right),\left(\rho',a'\right),t\right]=F\left(t\right)+G\left(t\right)\left(a+a'\right)+\frac{H\left(t\right)\left(\rho+\rho'\right)}{2}.$$

To determine $F\left(t\right)$, $G\left(t\right)$, and $H\left(t\right)$, I apply the envelope theorem to (6):

$$\begin{aligned}J_{ta}(\rho,a,t)=rJ_{a}(\rho,a,t)-u_{a}(\rho,a)-\alpha\int_{-1}^{1}[J_{a}(\rho',a,t)-J_{a}(\rho,a,t)]dF(\rho')\\-\lambda\int_{-\infty}^{\infty}\int_{-1}^{1}\left[\,J_{a}(\rho,a+q\left[\left(\rho,a\right),\left(\rho',a'\right)\right],t)-J_{a}(\rho,a,t)\right]\Phi(d\rho',da'|t).\end{aligned}\quad(7)$$

Using the conjectured marginal valuation and trade quantity:

$$\begin{aligned}F'\left(t\right)+2G'\left(t\right)a+H'\left(t\right)\rho=\left(r+\alpha+\lambda\right)\left(F\left(t\right)+2G\left(t\right)a+H\left(t\right)\rho\right)-\kappa_{0}+\kappa_{1}a+\kappa_{2}\rho\\-\alpha\left(F\left(t\right)+2G\left(t\right)a+H\left(t\right)\overline{\rho}\right)-\lambda\left(F\left(t\right)+2G\left(t\right)a+H\left(t\right)\rho\right)\\-\lambda\left(G\left(t\right)\left(A-a\right)+\frac{H\left(t\right)\left(\overline{\rho}-\rho\right)}{2}\right),\end{aligned}$$

where

$$\overline{\rho} \equiv \int_{-1}^{1} \rho dF(\rho)$$

and by market-clearing

$$\int_{-\infty}^{\infty} \int_{-1}^{1} a\Phi(d\rho, da|t) = A. \tag{8}$$

Coefficient matching implies that $F(t)$, $G(t)$, and $H(t)$ solve the following ODE's respectively:

$$\begin{aligned} F'(t) &= rF(t) - \kappa_0 - (\alpha + \lambda/2) H(t) \overline{\rho} - \lambda G(t) A, \\ G'(t) &= (r + \lambda/2) G(t) + \kappa_1/2, \\ H'(t) &= (r + \alpha + \lambda/2) H(t) + \kappa_2. \end{aligned}$$

General solutions of these ODE's are respectively:

$$\begin{aligned} F(t) &= \frac{\kappa_0}{r} - \frac{\kappa_1}{r} \frac{\lambda/2}{r + \lambda/2} A - \frac{\kappa_2}{r} \frac{\alpha + \lambda/2}{r + \alpha + \lambda/2} \overline{\rho} + c_F e^{r(t - T_R)}, \\ G(t) &= \frac{-\kappa_1}{2r + \lambda} + c_G e^{(r + \lambda/2)(t - T_R)}, \\ H(t) &= \frac{-\kappa_2}{r + \alpha + \lambda/2} + c_H e^{(r + \alpha + \lambda/2)(t - T_R)}. \end{aligned}$$

To find the constants c_F, c_G, and c_H, I use the boundary conditions in the limit as $t \to \infty$. This leads to the following coefficients for $t \geq T_R$:

$$\begin{aligned} F(t) &= \frac{\kappa_0}{r} - \frac{\kappa_1}{r} \frac{\lambda/2}{r + \lambda/2} A - \frac{\kappa_2}{r} \frac{\alpha + \lambda/2}{r + \alpha + \lambda/2} \overline{\rho}, \\ G(t) &= \frac{-\kappa_1}{2r + \lambda}, \\ H(t) &= \frac{-\kappa_2}{r + \alpha + \lambda/2}. \end{aligned}$$

At this point, we are able to characterize the terms of trade after the recovery.

Proposition 1 *In any equilibrium, investors' marginal valuations, individual trade sizes and transaction prices after the recovery (for $t \geq T_R$) are given by:*

$$J_a(\rho, a, t) = \frac{\kappa_0}{r} - \frac{\kappa_1}{r}\left[\frac{ra + (\lambda/2)\, A}{r + \lambda/2}\right] - \frac{\kappa_2}{r}\left[\frac{r\rho + (\alpha + \lambda/2)\overline{\rho}}{r + \alpha + \lambda/2}\right], \tag{9}$$

$$q\left[(\rho, a), (\rho', a'), t\right] = \frac{a' - a}{2} + \frac{\kappa_2}{\kappa_1}\frac{r + \lambda/2}{r + \alpha + \lambda/2}\frac{\rho' - \rho}{2} \tag{10}$$

and

$$\begin{aligned} P\left[(\rho, a), (\rho', a'), t\right] &= \frac{\kappa_0}{r} - \frac{\kappa_1}{r}\left[\frac{r(a + a')/2 + (\lambda/2)\, A}{r + \lambda/2}\right] \\ &\quad - \frac{\kappa_2}{r}\left[\frac{r(\rho + \rho')/2 + (\alpha + \lambda/2)\overline{\rho}}{r + \alpha + \lambda/2}\right]. \end{aligned} \tag{11}$$

After $D(t)$, $E(t)$, and $M(t)$ are found, the characterization of $J(\rho, a, t)$ after the recovery will be complete. I proceed by rewriting the conjectured $J(\rho, a, t)$:

$$J(\rho, a, t) = D(t) + E(t)\rho + Fa + Ga^2 + H\rho a + M(t)\rho^2.$$

Using the fact that $J(\rho, a, t)$ is quadratic in a, an exact second-order Taylor expansion shows that:

$$J(\rho, a + q, t) - J(\rho, a, t) = J_a(\rho, a + q, t)q - Gq^2.$$

Substituting into (6):

$$J_t(\rho, a, t) = (r + \alpha) J(\rho, a, t) - u(\rho, a) - \alpha \int_{-1}^{1} J(\rho', a, t) dF(\rho')$$
$$- \lambda \int_{-\infty}^{\infty} \int_{-1}^{1} \left\{ -2G \left(q \left[(\rho, a), (\rho', a'), t \right] \right)^2 \right\} \Phi(d\rho', da'|t).$$

Define

$$C \equiv \frac{H}{2G}.$$

Using the trade size function and the conjectured value function,

$$D'(t) + E'(t)\rho + M'(t)\rho^2 = (r + \alpha)\left(D(t) + E(t)\rho + Fa + Ga^2 + H\rho a + M(t)\rho^2\right)$$
$$- \kappa_0 a + \frac{1}{2}\kappa_1 a^2 + \kappa_2 \rho a - \alpha\left(D(t) + E(t)\overline{\rho} + Fa + Ga^2 + H\overline{\rho}a + M(t)\mathbb{E}\left[\rho^2\right]\right)$$
$$+ \frac{\lambda}{2} \int_{-\infty}^{\infty} \int_{-1}^{1} G\left(a' - a + C(\rho' - \rho)\right)^2 \Phi(d\rho', da'|t),$$

$$D'(t) + E'(t)\rho + M'(t)\rho^2 = (r + \alpha)\left(D(t) + E(t)\rho + Fa + Ga^2 + H\rho a + M(t)\rho^2\right)$$
$$- \kappa_0 a + \frac{1}{2}\kappa_1 a^2 + \kappa_2 \rho a - \alpha\left(D(t) + E(t)\overline{\rho} + Fa + Ga^2 + H\overline{\rho}a + M(t)\mathbb{E}\left[\rho^2\right]\right)$$
$$+ \frac{\lambda}{2} \left(GC^2\rho^2 - 2GC^2\rho\overline{\rho} + GC^2\mathbb{E}\left[\rho^2\right] + 2GCa\rho - 2GCa\overline{\rho} - 2GC\rho A + 2GC\mathbb{E}\left[\rho a|t\right]\right)$$
$$+ \frac{\lambda}{2} \left(G\mathbb{E}\left[a^2|t\right] - 2GaA + Ga^2\right).$$

Coefficient matching implies that $D(t)$, $E(t)$, and $M(t)$ solve the following ODE's respectively:

$$\begin{aligned}
D'(t) &= rD(t) - \alpha\left(E(t)\overline{\rho} + M(t)\mathbb{E}\left[\rho^2\right]\right) + \frac{\lambda}{2}G\,\mathbb{E}\left[(a+C\rho)^2\,|t\right], \\
E'(t) &= (r+\alpha)E(t) - \frac{\lambda}{2}H(A+C\overline{\rho}), \\
M'(t) &= (r+\alpha)M(t) + \frac{\lambda}{4}HC.
\end{aligned}$$

Therefore, to find the value function we need $\mathbb{E}\left[(a+C\rho)^2\,|t\right]$ which is determined in equilibrium. Define

$$\theta \equiv a + C\rho$$

as the *effective type* of the investor with asset holding a and correlation ρ. The following corollary shows the terms of trade as a function of investors' effective types.

Corollary 1 *In any equilibrium, the individual trade sizes and transaction prices after the recovery (for $t \geq T_R$) are given by:*

$$q(\theta, \theta', t) = \frac{\theta' - \theta}{2}$$

and

$$P(\theta, \theta', t) = \frac{\kappa_0}{r} - \frac{\kappa_1}{r}\frac{\lambda/2}{r+\lambda/2}A - \frac{\kappa_2}{r}\frac{\alpha+\lambda/2}{r+\alpha+\lambda/2}\overline{\rho} - \frac{\kappa_1}{r+\lambda/2}\left[\frac{\theta+\theta'}{2}\right],$$

where

$$\theta \equiv a + \frac{\kappa_2}{\kappa_1}\frac{r+\lambda/2}{r+\alpha+\lambda/2}\rho.$$

Corollary 1 shows that the sole determinant of the trade sizes is the difference between investors' effective types. This reveals an important intuition that C can be interpreted as a measure of how aggressively investors trade. When C is higher, investors' effective types will be more sensitive to their current correlation type. Thus, investors' asset position will fluctuate more as they will trade more aggressively, by putting more weight on their current correlation. When C is lower, the opposite will be true. Investors will put less weight on their current correlation, implying that they trade cautiously. This interpretation directly implies that investors trade more aggressively as frictions vanish (as $\lambda \to \infty$), since C is increasing in λ.

In addition, we know that, to find the equilibrium trading patterns and investors' values, knowing the distribution of θ is sufficient. In the next subsection, I will analyze the equilibrium dynamics for the distribution of θ.

3.1.3 Dynamics of the distribution of investors' states

For simplicity, I assume that the equilibrium conditional distribution of effective types have densities. This assumption is actually not necessary but simplifies the presentation of the results. Suppose $g(\theta|t)$ and $g_\rho(\theta|t)$ are the unconditional and conditional pdf of θ, respectively, for $t \geq T_R$. n-th conditional moment of the effective types can be written as follows using the Fourier transform (Bracewell, 2000):

$$\mathbb{E}\left[\theta^n \mid \rho, t\right] = (-i2\pi)^{-n} \left[\frac{d^n}{dz^n}\widehat{g}_\rho(z|t)\right]_{z=0}.$$

Since I have explicit expression for trade sizes, I can derive the law of motion for the investors' effective types following the same steps in the derivation of steady-state conditions in Chapter 1. The resulting law of motion is:

$$\frac{d}{dt}g_\rho(\theta|t) = -(\alpha+2\lambda)g_\rho(\theta|t) + \alpha\int_{-1}^{1} g_{\rho'}(\theta + C(\rho'-\rho)|t)dF(\rho')$$
$$+4\lambda\int_{-1}^{1}\int_{-\infty}^{\infty} g_\rho(\theta'|t)g_{\rho'}(2\theta-\theta'|t)d\theta' dF(\rho').$$

This equation has the usual inflow-outflow interpretation. The first term represents the outflow due to idiosyncratic shocks and trade. The second and third terms represent the inflow due to idiosyncratic shocks and the inflow due to trade, respectively. The last term contains a convolution since any investor with type (ρ,θ') can become of type (ρ,θ) if she meets the right counterparty. The right counterparty in this context means an investor with type $(\rho', 2\theta-\theta')$. Corollary 1 immediately implies that the post-trade effective type of the first investor will be θ, and hence she will create inflow. Since the convolution term complicates the computation of distribution function, I will make use of the Fourier transform of this law of motion, which is:

$$\frac{d}{dt}\widehat{g}_\rho(z|t) = -(\alpha+2\lambda)\widehat{g}_\rho(z|t) + \alpha\int_{-1}^{1} e^{i2\pi C(\rho'-\rho)z}\widehat{g}_{\rho'}(z|t)dF(\rho') + 2\lambda\int_{-1}^{1}\widehat{g}_\rho(\frac{z}{2}|t)\widehat{g}_{\rho'}(\frac{z}{2}|t)dF(\rho').$$

Then I use the Fourier transform of θ distribution to find an expression for $\frac{d^n}{dz^n}\widehat{g}_\rho(z|t)$:

$$\begin{aligned}\frac{d}{dt}\left(\frac{d^n}{dz^n}\widehat{g}_\rho(z|t)\right) &= -(\alpha+2\lambda)\frac{d^n}{dz^n}\widehat{g}_\rho(z|t) + \alpha\int_{-1}^{1}\frac{d^n}{dz^n}\left(e^{i2\pi C(\rho'-\rho)z}\widehat{g}_{\rho'}(z|t)\right)dF(\rho') \\ &\quad +2\lambda\int_{-1}^{1}\frac{d^n}{dz^n}\left(\widehat{g}_\rho(\frac{z}{2}|t)\widehat{g}_{\rho'}(\frac{z}{2}|t)\right)dF(\rho').\end{aligned}$$

To proceed, I use the following generalization of the product rule:

$$\frac{d^n}{dx^n}\prod_{i=1}^{2} g_i(x) = \sum_{j_1+j_2=n} \begin{pmatrix} n \\ j_1, j_2 \end{pmatrix} \prod_{i=1}^{2} \frac{d^{j_i}}{dx^{j_i}} g_i(x),$$

$$\begin{aligned}
\frac{d}{dt}\left(\frac{d^n}{dz^n}\widehat{g}_\rho(z|t)\right) &= -(\alpha+2\lambda)\frac{d^n}{dz^n}\widehat{g}_\rho(z|t) \\
+\alpha\int_{-1}^{1}\sum_{j_1+j_2=n} &\begin{pmatrix} n \\ j_1, j_2 \end{pmatrix} \left\{\left[\frac{d^{j_1}}{dz^{j_1}}e^{i2\pi C(\rho'-\rho)z}\right]\left[\frac{d^{j_2}}{dz^{j_2}}\widehat{g}_{\rho'}(z|t)\right]\right\} dF(\rho') \\
+2\lambda\int_{-1}^{1}\sum_{j_1+j_2=n} &\begin{pmatrix} n \\ j_1, j_2 \end{pmatrix} \left\{\left[\frac{d^{j_1}}{dz^{j_1}}\widehat{g}_\rho(\frac{z}{2}|t)\right]\left[\frac{d^{j_2}}{dz^{j_2}}\widehat{g}_{\rho'}(\frac{z}{2}|t)\right]\right\} dF(\rho'),
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt}\widehat{g}^{(n)}_\rho(z|t) &= -(\alpha+2\lambda)\widehat{g}^{(n)}_\rho(z|t) \\
+\alpha\int_{-1}^{1}\sum_{j_1+j_2=n} &\begin{pmatrix} n \\ j_1, j_2 \end{pmatrix} \left\{(i2\pi C(\rho'-\rho))^{j_1}e^{i2\pi C(\rho'-\rho)z}\widehat{g}^{(j_2)}_{\rho'}(z|t)\right\} dF(\rho') \\
+2\lambda\int_{-1}^{1}\sum_{j_1+j_2=n} &\begin{pmatrix} n \\ j_1, j_2 \end{pmatrix} (\frac{1}{2})^n \widehat{g}^{(j_1)}_\rho(\frac{z}{2}|t)\widehat{g}^{(j_2)}_{\rho'}(\frac{z}{2}|t) dF(\rho'),
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt}\widehat{g}^{(n)}_\rho(0|t) = -(\alpha+2\lambda)\widehat{g}^{(n)}_\rho(0|t) + \alpha\int_{-1}^{1}\sum_{j_1+j_2=n} &\begin{pmatrix} n \\ j_1, j_2 \end{pmatrix} \left\{(i2\pi C(\rho'-\rho))^{j_1}\widehat{g}^{(j_2)}_{\rho'}(0|t)\right\} dF(\rho') \\
+2\lambda\int_{-1}^{1}\sum_{j_1+j_2=n} &\begin{pmatrix} n \\ j_1, j_2 \end{pmatrix} (\frac{1}{2})^n \widehat{g}^{(j_1)}_\rho(0|t)\widehat{g}^{(j_2)}_{\rho'}(0|t) dF(\rho').
\end{aligned}$$

Dividing both sides by $(-i2\pi)^n$:

$$\frac{d}{dt}\mathbb{E}\left[\theta^n \mid \rho, t\right] = -(\alpha+2\lambda)\mathbb{E}\left[\theta^n \mid \rho, t\right]$$
$$+\alpha\int_{-1}^{1}\sum_{j_1+j_2=n}\binom{n}{j_1, j_2}\left\{(-C(\rho'-\rho))^{j_1}\mathbb{E}\left[\theta^{j_2} \mid \rho', t\right]\right\}dF(\rho')$$
$$+2\lambda(\frac{1}{2})^n\int_{-1}^{1}\sum_{j_1+j_2=n}\binom{n}{j_1, j_2}\mathbb{E}\left[\theta^{j_1} \mid \rho, t\right]\mathbb{E}\left[\theta^{j_2} \mid \rho', t\right]dF(\rho').$$

Using the binomial expansion of $(-C(\rho-\rho'))^{j_1}$:

$$\frac{d}{dt}\mathbb{E}\left[\theta^n \mid \rho, t\right] = -(\alpha+2\lambda)\mathbb{E}\left[\theta^n \mid \rho, t\right]$$
$$+\alpha\int_{-1}^{1}\sum_{j_1+j_2=n}\binom{n}{j_1, j_2}\left\{C^{j_1}\sum_{k=0}^{j_1}\binom{j_1}{k}(-\rho')^k(\rho)^{j_1-k}\mathbb{E}\left[\theta^{j_2} \mid \rho', t\right]\right\}dF(\rho')$$
$$+2\lambda(\frac{1}{2})^n\int_{-1}^{1}\sum_{j_1+j_2=n}\binom{n}{j_1, j_2}\mathbb{E}\left[\theta^{j_1} \mid \rho, t\right]\mathbb{E}\left[\theta^{j_2} \mid \rho', t\right]dF(\rho'),$$

$$\frac{d}{dt}\mathbb{E}\left[\theta^n \mid \rho, t\right] = -(\alpha+2\lambda)\mathbb{E}\left[\theta^n \mid \rho, t\right]$$
$$+\alpha\sum_{j_1+j_2=n}\binom{n}{j_1, j_2}C^{j_1}\sum_{k=0}^{j_1}\binom{j_1}{k}(\rho)^{j_1-k}\int_{-1}^{1}(-\rho')^k\mathbb{E}\left[\theta^{j_2} \mid \rho', t\right]dF(\rho')$$
$$+2\lambda(\frac{1}{2})^n\sum_{j_1+j_2=n}\binom{n}{j_1, j_2}\mathbb{E}\left[\theta^{j_1} \mid \rho, t\right]\int_{-1}^{1}\mathbb{E}\left[\theta^{j_2} \mid \rho', t\right]dF(\rho').$$

Applying the law of iterated expectations, I arrive at the following proposition.

Proposition 2 *Let* $C = \frac{\kappa_2}{\kappa_1}\frac{r+\lambda/2}{r+\alpha+\lambda/2}$*. The following system of ODEs characterizes all mo-*

ments of the equilibrium conditional distributions of effective type θ after the recovery (for $t \geq T_R$):

$$\frac{d}{dt}\mathbb{E}\left[\theta^n \mid \rho, t\right] = -(\alpha + 2\lambda)\mathbb{E}\left[\theta^n \mid \rho, t\right] + 2\lambda(\frac{1}{2})^n \sum_{j_1+j_2=n} \begin{pmatrix} n \\ j_1, j_2 \end{pmatrix} \mathbb{E}\left[\theta^{j_1} \mid \rho, t\right] \mathbb{E}\left[\theta^{j_2}|t\right]$$

$$+\alpha \sum_{j_1+j_2=n} \begin{pmatrix} n \\ j_1, j_2 \end{pmatrix} C^{j_1} \sum_{k=0}^{j_1} \begin{pmatrix} j_1 \\ k \end{pmatrix} (-1)^k (\rho)^{j_1-k} \mathbb{E}\left[\theta^{j_2}\rho^k|t\right]$$

for all $\rho \in supp(F)$ and for all $n \in \mathbb{Z}_+$; and

$$\mathbb{E}\left[\theta|t\right] = A + C\overline{\rho}.$$

The following corollary shows some relevant moments in closed form.

Corollary 2 *Let $C = \frac{\kappa_2}{\kappa_1}\frac{r+\lambda/2}{r+\alpha+\lambda/2}$. For $t \geq T_R$,*

$$\mathbb{E}\left[\theta \mid \rho, t\right] = \left[1 - e^{-(\alpha+\lambda)(t-T_R)}\right]\left(A + C\frac{\alpha\rho + \lambda\overline{\rho}}{\alpha+\lambda}\right) + e^{-(\alpha+\lambda)(t-T_R)}\mathbb{E}\left[\theta \mid \rho, T_R\right],$$

$$\mathbb{E}\left[\theta^2 \mid t\right] = \left[1 - e^{-\lambda(t-T_R)}\right]\left[\frac{2\alpha}{\alpha+\lambda}C^2 var\left(\rho\right) + \left(A + C\overline{\rho}\right)^2\right] + e^{-\lambda(t-T_R)}\mathbb{E}\left[\theta^2 \mid T_R\right],$$

and

$$var\left[\theta \mid t\right] = \left[1 - e^{-\lambda(t-T_R)}\right]\frac{2\alpha}{\alpha+\lambda}C^2 var\left(\rho\right) + e^{-\lambda(t-T_R)} var\left[\theta \mid T_R\right].$$

Now we can go back to the characterization of the remaining coefficients of the value function $J(\rho, a, t)$.

$$\begin{aligned} D'(t) &= rD(t) - \alpha\left(E(t)\overline{\rho} + M(t)\mathbb{E}\left[\rho^2\right]\right) + \frac{\lambda}{2}G\,\mathbb{E}\left[(a + C\rho)^2 \,|t\right], \\ E'(t) &= (r+\alpha)E(t) - \frac{\lambda}{2}H(A + C\overline{\rho}), \\ M'(t) &= (r+\alpha)M(t) + \frac{\lambda}{4}HC. \end{aligned}$$

Solving the ODEs and using Corollary 2,

$$\begin{aligned} D(t) &= \frac{\alpha\left(E\overline{\rho} + M\mathbb{E}\left[\rho^2\right]\right)}{r} - \frac{\lambda}{2}\frac{G}{r}\left[\frac{2\alpha}{\alpha+\lambda}C^2 var(\rho) + (A + C\overline{\rho})^2\right] \\ &\quad -\frac{\lambda}{2}\frac{G}{r+\lambda}\left[\mathbb{E}\left[\theta^2 \mid T_R\right] - \frac{2\alpha}{\alpha+\lambda}C^2 var(\rho) - (A + C\overline{\rho})^2\right]e^{-\lambda(t-T_R)}, \\ E(t) &= -\frac{\lambda(A + C\overline{\rho})}{2(r+\alpha)}\frac{\kappa_2}{r+\alpha+\lambda/2}, \\ M(t) &= \frac{\lambda}{4(r+\alpha)}\frac{(\kappa_2)^2}{\kappa_1}\frac{r+\lambda/2}{(r+\alpha+\lambda/2)^2}. \end{aligned}$$

3.2 The crisis equilibrium

Now, let me move on to describing the equilibrium path during the crisis, i.e. when $t \in [0, T_R)$. Suppose $V(\rho, a, t)$ denotes the maximum attainable continuation utility of an investor with type ρ and asset holding a at date t. Taking the pricing function $P[(\rho, a), (\rho', a'), t]$ and the trade size function $q[(\rho, a), (\rho', a'), t]$ as given, $V(\rho, a, t)$ is defined as

$$V(\rho, a, t) \equiv \mathbb{E}_t\left[\int_t^{T_R} e^{-r(s-t)}\left[\overline{u}(\rho_s, a_s) + c_s\right]ds + e^{-r(T_R-t)}J(\rho, a, T_R) \mid \rho_t = \rho,\ a_t = a\right],$$

s.t.

$$da_t = \begin{cases} q\left[\left(\rho_t, a_t\right), \left(\rho_t', a_t'\right), t\right] & \text{if there is contact with investor } \left(\rho_t', a_t'\right) \\ 0 & \text{if no contact,} \end{cases}$$

$$c_t = \begin{cases} -P\left[\left(\rho_t, a_t\right), \left(\rho_t', a_t'\right), t\right] da_t & \text{if there is contact with investor } \left(\rho_t', a_t'\right) \\ 0 & \text{if no contact.} \end{cases}$$

Taking the derivative with respect to t and rearranging, one can show that $V(\rho, a, t)$ satisfies the following ordinary differential equation (ODE):

$$\begin{aligned} V_t(\rho, a, t) = rV(\rho, a, t) - \overline{u}(\rho, a) - \alpha \int_{-1}^{1} [V(\rho', a, t) - V(\rho, a, t)] dF(\rho') \\ - 2\lambda \int_{-\infty}^{\infty} \int_{-1}^{1} \{V(\rho, a + q\left[(\rho, a), (\rho', a'), t\right], t) - V(\rho, a, t) \\ -q\left[(\rho, a), (\rho', a'), t\right] P\left[(\rho, a), (\rho', a'), t\right]\} \Phi(d\rho', da'|t) \\ - R[J(\rho, a, T_R|t) - V(\rho, a, t)]. \quad (12) \end{aligned}$$

Terms of trade, $q\left[(\rho, a), (\rho', a'), t\right]$ and $P\left[(\rho, a), (\rho', a'), t\right]$, solve the same Nash bargaining problem (3) in which the function J is replaced with V. Substituting the trade quantity and price into (12), I get

$$\begin{aligned} V_t(\rho, a, t) = rV(\rho, a, t) - \overline{u}(\rho, a) - \alpha \int_{-1}^{1} [V(\rho', a, t) - V(\rho, a, t)] dF(\rho') \\ - \lambda \int_{-\infty}^{\infty} \int_{-1}^{1} \left[\max_q \ \{V(\rho, a + q, t) - V(\rho, a, t) + V(\rho', a' - q, t) - V(\rho', a', t)\}\right] \Phi(d\rho', da'|t) \\ - R[J(\rho, a, T_R|t) - V(\rho, a, t)]. \quad (13) \end{aligned}$$

To solve for $V(\rho, a, t)$, I again resort to a guess&verify approach. My conjecture is

$$V(\rho, a, t) = \overline{D}(t) + \overline{E}(t)\rho + \overline{F}(t)a + \overline{G}(t)a^2 + \overline{H}(t)\rho a + \overline{M}(t)\rho^2.$$

The solution of Nash-bargaining implies that

$$q[(\rho, a), (\rho', a'), t] = \frac{a' - a}{2} + \frac{\overline{H}(t)(\rho' - \rho)}{4\overline{G}(t)}$$

and

$$P[(\rho, a), (\rho', a'), t] = \overline{F}(t) + \overline{G}(t)(a + a') + \frac{\overline{H}(t)(\rho + \rho')}{2}.$$

To determine $\overline{F}(t)$, $\overline{G}(t)$, and $\overline{H}(t)$, I apply the envelope theorem to (13):

$$\begin{aligned} V_{ta}(\rho, a, t) = rV_a(\rho, a, t) - \overline{u}_a(\rho, a) - \alpha \int_{-1}^{1} [V_a(\rho', a, t) - V_a(\rho, a, t)] dF(\rho') \\ - \lambda \int_{-\infty}^{\infty} \int_{-1}^{1} [\, V_a(\rho, a + q[(\rho, a), (\rho', a')], t) - V_a(\rho, a, t)] \, \Phi(d\rho', da'|t) \\ - R[J_a(\rho, a, T_R) - V_a(\rho, a, t)]. \end{aligned} \tag{14}$$

Using the conjectured marginal valuation, trade quantity, and the solution of J_a from Proposition 1:

$$\begin{aligned} \overline{F}'(t) + 2\overline{G}'(t)a + \overline{H}'(t)\rho = (r + \alpha + \lambda/2 + R)\left(\overline{F}(t) + 2\overline{G}(t)a + \overline{H}(t)\rho\right) - \kappa_0 + \overline{\kappa}_1 a + \overline{\kappa}_2 \rho \\ - \alpha\left(\overline{F}(t) + 2\overline{G}(t)a + \overline{H}(t)\overline{\rho}\right) - \lambda\left(\overline{F}(t) + \overline{G}(t)A + \overline{H}(t)\frac{\rho}{2}\right) \\ - R\left(\frac{\kappa_0}{r} - \frac{\kappa_1}{r}\frac{ra + (\lambda/2)A}{r + \lambda/2} - \frac{\kappa_2}{r}\frac{r\rho + (\alpha + \lambda/2)\overline{\rho}}{r + \alpha + \lambda/2}\right). \end{aligned}$$

Coefficient matching implies that $\overline{F}(t)$, $\overline{G}(t)$, and $\overline{H}(t)$ solve the following ODE's respectively:

$$\begin{aligned}
\overline{F}'(t) &= (r+R)\overline{F}(t) - \kappa_0 - (\alpha+\lambda/2)\overline{H}(t)\overline{\rho} - \lambda\overline{G}(t)A \\
&\quad -R\left(\frac{\kappa_0}{r} - \frac{\kappa_1}{r}\frac{(\lambda/2)A}{r+\lambda/2} - \frac{\kappa_2}{r}\frac{(\alpha+\lambda/2)\overline{\rho}}{r+\alpha+\lambda/2}\right), \\
\overline{G}'(t) &= (r+\lambda/2+R)\overline{G}(t) + \overline{\kappa}_1/2 + R\frac{\kappa_1}{2r+\lambda}, \\
\overline{H}'(t) &= (r+\alpha+\lambda/2+R)\overline{H}(t) + \overline{\kappa}_2 + R\frac{\kappa_2}{r+\alpha+\lambda/2}.
\end{aligned}$$

General solutions of these ODE's are respectively (using the limiting condition as $R \to 0$ as the boundary condition):

$$\begin{aligned}
\overline{F}(t) &= \frac{\kappa_0}{r} - \frac{1}{r}\frac{r\widetilde{\kappa}_1 + R\kappa_1}{r+R}\frac{(\lambda/2)A}{r+\lambda/2} - \frac{1}{r}\frac{r\widetilde{\kappa}_2 + R\kappa_2}{r+R}\frac{(\alpha+\lambda/2)\overline{\rho}}{r+\alpha+\lambda/2}, \\
\overline{G}(t) &= \frac{-\widetilde{\kappa}_1}{2r+\lambda}, \\
\overline{H}(t) &= \frac{-\widetilde{\kappa}_2}{r+\alpha+\lambda/2},
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{\kappa}_1 &\equiv \frac{(r+\lambda/2)\overline{\kappa}_1 + R\kappa_1}{r+\lambda/2+R}, \\
\widetilde{\kappa}_2 &\equiv \frac{(r+\alpha+\lambda/2)\overline{\kappa}_2 + R\kappa_2}{r+\alpha+\lambda/2+R}.
\end{aligned}$$

At this point, we are able to characterize the terms of trade during the crisis.

Proposition 3 *Let* $\widetilde{\kappa}_1 = \frac{(r+\lambda/2)\overline{\kappa}_1 + R\kappa_1}{r+\lambda/2+R}$ *and* $\widetilde{\kappa}_2 = \frac{(r+\alpha+\lambda/2)\overline{\kappa}_2 + R\kappa_2}{r+\alpha+\lambda/2+R}$. *In any equilibrium, investors' marginal valuations, individual trade sizes and transaction prices during the* *crisis (for $t < T_R$) are given by:*

$$
\begin{aligned}
V_a(\rho,a,t) &= \frac{\kappa_0}{r} - \frac{1}{r}\frac{r\widetilde{\kappa}_1 + R\kappa_1}{r+R}\frac{(\lambda/2)\,A}{r+\lambda/2} - \frac{1}{r}\frac{r\widetilde{\kappa}_2 + R\kappa_2}{r+R}\frac{(\alpha+\lambda/2)\overline{\rho}}{r+\alpha+\lambda/2} \\
&\quad - \frac{\widetilde{\kappa}_1}{r+\lambda/2}a - \frac{\widetilde{\kappa}_2}{r+\alpha+\lambda/2}\rho, \qquad (19)
\end{aligned}
$$

$$
q\left[(\rho,a),(\rho',a'),t\right] = \frac{a'-a}{2} + \frac{\widetilde{\kappa}_2}{\widetilde{\kappa}_1}\frac{r+\lambda/2}{r+\alpha+\lambda/2}\frac{\rho'-\rho}{2}, \qquad (20)
$$

and

$$
\begin{aligned}
P\left[(\rho,a),(\rho',a'),t\right] &= \frac{\kappa_0}{r} - \frac{1}{r}\frac{r\widetilde{\kappa}_1 + R\kappa_1}{r+R}\frac{(\lambda/2)\,A}{r+\lambda/2} - \frac{1}{r}\frac{r\widetilde{\kappa}_2 + R\kappa_2}{r+R}\frac{(\alpha+\lambda/2)\overline{\rho}}{r+\alpha+\lambda/2} \\
&\quad - \frac{\widetilde{\kappa}_1}{r+\lambda/2}\left(\frac{a+a'}{2}\right) - \frac{\widetilde{\kappa}_2}{r+\alpha+\lambda/2}\left(\frac{\rho+\rho'}{2}\right). \qquad (21)
\end{aligned}
$$

After $\overline{D}(t)$, $\overline{E}(t)$, and $\overline{M}(t)$ are found, the characterization of $V(\rho,a,t)$ after the recovery will be complete. I proceed by rewriting the conjectured $V(\rho,a,t)$:

$$
V(\rho,a,t) = \overline{D}(t) + \overline{E}(t)\,\rho + \overline{F}a + \overline{G}a^2 + \overline{H}\rho a + \overline{M}(t)\,\rho^2.
$$

Using the fact that $V(\rho,a,t)$ is quadratic in a, an exact second-order Taylor expansion shows that:

$$
V(\rho,a+q,t) - V(\rho,a,t) = V_a(\rho,a+q,t)q - \overline{G}q^2.
$$

Substituting into (13):

$$
\begin{aligned}
V_t(\rho,a,t) &= (r+\alpha+R)\,V(\rho,a,t) - \overline{u}(\rho,a) - \alpha\int_{-1}^{1} V(\rho',a,t)dF(\rho') \\
&\quad - \lambda\int_{-\infty}^{\infty}\int_{-1}^{1}\left\{-2\overline{G}\left(q\left[(\rho,a),(\rho',a'),t\right]\right)^2\right\}\Phi(d\rho',da'|t) - RJ(\rho,a,T_R|t).
\end{aligned}
$$

Define

$$\overline{C} \equiv \frac{\overline{H}}{2\overline{G}}.$$

Using the trade size function, the conjectured value function, and the solution of J,

$$\begin{aligned}
\overline{D}'(t) + \overline{E}'(t)\rho + \overline{M}'(t)\rho^2 &= (r+\alpha+R)\left(\overline{D}(t) + \overline{E}(t)\rho + \overline{F}a + \overline{G}a^2 + \overline{H}\rho a + \overline{M}(t)\rho^2\right) \\
&- \kappa_0 a + \frac{1}{2}\overline{\kappa}_1 a^2 + \overline{\kappa}_2\rho a - \alpha\left(\overline{D}(t) + \overline{E}(t)\overline{\rho} + \overline{F}a + \overline{G}a^2 + \overline{H}\overline{\rho}a + \overline{M}(t)\mathbb{E}\left[\rho^2\right]\right) \\
+\frac{\lambda}{2}\int_{-\infty}^{\infty}\int_{-1}^{1} &\overline{G}\left(a' - a + \overline{C}(\rho'-\rho)\right)^2 \Phi(d\rho', da'|t) - R\left(D\left(T_R\right) + E\rho + Fa + Ga^2 + H\rho a + M\rho^2\right),
\end{aligned}$$

$$\begin{aligned}
\overline{D}'(t) + \overline{E}'(t)\rho + \overline{M}'(t)\rho^2 &= (r+\alpha+R)\left(\overline{D}(t) + \overline{E}(t)\rho + \overline{F}a + \overline{G}a^2 + \overline{H}\rho a + \overline{M}(t)\rho^2\right) \\
&- \kappa_0 a + \frac{1}{2}\overline{\kappa}_1 a^2 + \overline{\kappa}_2\rho a - \alpha\left(\overline{D}(t) + \overline{E}(t)\overline{\rho} + \overline{F}a + \overline{G}a^2 + \overline{H}\overline{\rho}a + \overline{M}(t)\mathbb{E}\left[\rho^2\right]\right) \\
&+ \frac{\lambda}{2}\left(\overline{GC}^2\rho^2 - 2\overline{GC}^2\rho\overline{\rho} + \overline{GC}^2\mathbb{E}\left[\rho^2\right] + 2\overline{GC}a\rho - 2\overline{GC}a\overline{\rho} - 2\overline{GC}\rho A + 2\overline{GC}\mathbb{E}\left[\rho a|t\right]\right) \\
&+ \frac{\lambda}{2}\left(\overline{G}\mathbb{E}\left[a^2|t\right] - 2\overline{G}aA + \overline{G}a^2\right) - R\left(D\left(T_R|t\right) + E\rho + Fa + Ga^2 + H\rho a + M\rho^2\right).
\end{aligned}$$

Coefficient matching implies that $\overline{D}(t)$, $\overline{E}(t)$, and $\overline{M}(t)$ solve the following ODE's respectively:

$$\begin{aligned}
\overline{D}'(t) &= (r+R)\overline{D}(t) - \alpha\left(\overline{E}(t)\overline{\rho} + \overline{M}(t)\mathbb{E}\left[\rho^2\right]\right) + \frac{\lambda}{2}\overline{G}\,\mathbb{E}\left[\left(a + \overline{C}\rho\right)^2|t\right] - RD\left(T_R|t\right), \\
\overline{E}'(t) &= (r+\alpha+R)\overline{E}(t) - \frac{\lambda}{2}\overline{H}\left(A + \overline{C}\overline{\rho}\right) - RE, \\
\overline{M}'(t) &= (r+\alpha+R)\overline{M}(t) + \frac{\lambda}{4}\overline{HC} - RM.
\end{aligned}$$

Solving the ODEs and using Corollary 2,

$$\begin{aligned}
\overline{D}(t) &= \frac{\alpha\left(E\overline{\rho}+M\mathbb{E}\left[\rho^2\right]\right)}{r} - \frac{\lambda}{2}\frac{G}{r}\left[\frac{2\alpha}{\alpha+\lambda}C^2 var(\rho) + (A+C\overline{\rho})^2\right] \\
&\quad -\frac{\lambda}{2}\frac{G}{r+\lambda}\left[\mathbb{E}\left[\mathbb{E}\left[\theta^2 \mid T_R\right]\right] - \frac{2\alpha}{\alpha+\lambda}C^2 var(\rho) - (A+C\overline{\rho})^2\right], \\
\overline{E}(t) &= -\frac{1}{r+\alpha}\frac{\lambda}{2}\frac{1}{(r+\alpha+\lambda/2)^2}\left(\frac{r+\alpha}{r+\alpha+R}\left(A+\overline{C}\overline{\rho}\right)\widetilde{\kappa}_2 + \frac{R}{r+\alpha+R}\left(A+C\overline{\rho}\right)\kappa_2\right), \\
\overline{M}(t) &= \frac{1}{r+\alpha}\frac{\lambda}{4}\frac{r+\lambda/2}{(r+\alpha+\lambda/2)^2}\left(\frac{r+\alpha}{r+\alpha+R}\frac{(\widetilde{\kappa}_2)^2}{\widetilde{\kappa}_1} + \frac{R}{r+\alpha+R}\frac{(\kappa_2)^2}{\kappa_1}\right).
\end{aligned}$$

Again the relevant equilibrium objects are the functions of investors' effective types defined as

$$\overline{\theta} \equiv a + \overline{C}\rho.$$

The following corollary shows the terms of trade as a function of investors' effective types.

Corollary 3 *Let* $\widetilde{\kappa}_1 = \frac{(r+\lambda/2)\overline{\kappa}_1 + R\kappa_1}{r+\lambda/2+R}$ *and* $\widetilde{\kappa}_2 = \frac{(r+\alpha+\lambda/2)\overline{\kappa}_2 + R\kappa_2}{r+\alpha+\lambda/2+R}$. *In any equilibrium, the individual trade sizes and transaction prices during the crisis (for* $t < T_R$*) are given by:*

$$q(\overline{\theta},\overline{\theta}',t) = \frac{\overline{\theta}'-\overline{\theta}}{2}$$

and

$$\begin{aligned}
P(\overline{\theta},\overline{\theta}',t) &= \frac{\kappa_0}{r} - \frac{1}{r}\frac{r\widetilde{\kappa}_1 + R\kappa_1}{r+R}\frac{(\lambda/2)A}{r+\lambda/2} - \frac{1}{r}\frac{r\widetilde{\kappa}_2 + R\kappa_2}{r+R}\frac{(\alpha+\lambda/2)\overline{\rho}}{r+\alpha+\lambda/2} \\
&\quad -\frac{\widetilde{\kappa}_1}{r+\lambda/2}\left(\frac{\overline{\theta}+\overline{\theta}'}{2}\right),
\end{aligned}$$

where

$$\overline{\theta} \equiv a + \frac{\widetilde{\kappa}_2}{\widetilde{\kappa}_1}\frac{r+\lambda/2}{r+\alpha+\lambda/2}\rho.$$

Since the crisis trade quantity as a function of effective types is the same as the trade quantity function after recovery, Proposition 2 apply to the crisis equilibrium as well after a relabeling of θ and C with bars. Thus, I derive the following corollary to show some relevant moments in closed form.

Corollary 4 *Let* $\widetilde{\kappa}_1 = \frac{(r+\lambda/2)\overline{\kappa}_1 + R\kappa_1}{r+\lambda/2+R}$, $\widetilde{\kappa}_2 = \frac{(r+\alpha+\lambda/2)\overline{\kappa}_2 + R\kappa_2}{r+\alpha+\lambda/2+R}$, and $\overline{\theta} \equiv a + \frac{\widetilde{\kappa}_2}{\widetilde{\kappa}_1}\frac{r+\lambda/2}{r+\alpha+\lambda/2}\rho$. *For* $t < T_R$,

$$\mathbb{E}\left[\overline{\theta} \mid \rho, t\right] = \left[1 - e^{-(\alpha+\lambda)t}\right]\left(A + \overline{C}\frac{\alpha\rho + \lambda\overline{\rho}}{\alpha+\lambda}\right) + e^{-(\alpha+\lambda)t}\mathbb{E}\left[\overline{\theta} \mid \rho, 0\right],$$

$$\mathbb{E}\left[\overline{\theta}^2 \mid t\right] = \left[1 - e^{-\lambda t}\right]\left[\frac{2\alpha}{\alpha+\lambda}\overline{C}^2 var\left(\rho\right) + \left(A + \overline{C}\overline{\rho}\right)^2\right] + e^{-\lambda t}\mathbb{E}\left[\overline{\theta}^2 \mid 0\right],$$

and

$$var\left[\overline{\theta} \mid t\right] = \left[1 - e^{-\lambda t}\right]\frac{2\alpha}{\alpha+\lambda}\overline{C}^2 var\left(\rho\right) + e^{-\lambda t} var\left[\overline{\theta} \mid 0\right],$$

where

$$\mathbb{E}\left[\overline{\theta} \mid \rho, 0\right] = A + \frac{\lambda}{\alpha+\lambda}C\left(\overline{\rho} - \rho\right) + \overline{C}\rho,$$

$$\mathbb{E}\left[\overline{\theta}^2 \mid 0\right] = \left(C^2 + \overline{C}^2 - 2C\overline{C}\frac{\lambda}{\alpha+\lambda}\right) var\left(\rho\right) + \left(A + \overline{C}\overline{\rho}\right)^2,$$

and

$$var\left[\overline{\theta} \mid 0\right] = \left(C^2 + \overline{C}^2 - 2C\overline{C}\frac{\lambda}{\alpha+\lambda}\right) var\left(\rho\right).$$

4 Results

4.1 Price dispersion

As the measure of price dispersion, I use the standard deviation σ_P of the equilibrium price distribution. Following Corollary 1 and, and using the fact that θ and θ' are i.i.d. due to random matching,

$$\sigma_P(t) = \begin{cases} \frac{\widetilde{\kappa}_1}{\sqrt{2}(r+\lambda/2)}\sigma_{\overline{\theta}}(t) & \text{for } t < T_R \\ \frac{\kappa_1}{\sqrt{2}(r+\lambda/2)}\sigma_{\theta}(t) & \text{for } t \geq T_R, \end{cases}$$

where $\sigma_{\overline{\theta}}(t)$ and $\sigma_{\overline{\theta}}(t)$ are the standard deviation of effective types before and after recovery, respectively. $1/\sqrt{2}(=\sqrt{2}\sqrt{\frac{1}{4}})$ is due to the fact that the trade is bilateral, and the transaction price is affected by both counterparties' effective types; $\frac{\widetilde{\kappa}_1}{(r+\lambda/2)}$ and $\frac{\kappa_1}{(r+\lambda/2)}$ are the sensitivity of price to effective type before and after recovery, respectively. An expression for $\sigma_{\overline{\theta}}(t)$ and $\sigma_{\theta}(t)$ can easily be derived from Corollary 2 and 4:

$$\begin{aligned} \sigma_{\overline{\theta}}(t) &= \sqrt{\left[1-e^{-\lambda t}\right]\tfrac{2\alpha}{\alpha+\lambda}\overline{C}^2 var(\rho) + e^{-\lambda t}\left(C^2+\overline{C}^2-2C\overline{C}\tfrac{\lambda}{\alpha+\lambda}\right)var(\rho)} \quad \text{for } t < T_R \\ \sigma_{\theta}(t) &= \sqrt{\left[1-e^{-\lambda(t-T_R)}\right]\tfrac{2\alpha}{\alpha+\lambda}C^2 var(\rho) + e^{-\lambda(t-T_R)}var\left[\theta \mid T_R\right]} \quad \text{for } t \geq T_R, \end{aligned}$$

where

$$\begin{aligned} var\left[\theta \mid T_R\right] &= \left[1-e^{-\lambda T_R}\right]\frac{2\alpha}{\alpha+\lambda}\overline{C}^2 var(\rho) + e^{-\lambda T_R}\left(C^2+\overline{C}^2-2C\overline{C}\frac{\lambda}{\alpha+\lambda}\right)var(\rho) \\ &\quad + \left(C^2-\overline{C}^2\right)var(\rho) - \left[1-e^{-(\alpha+\lambda)T_R}\right]2\left(C-\overline{C}\right)\overline{C}\frac{\lambda}{\alpha+\lambda}var(\rho) \\ &\quad - e^{-(\alpha+\lambda)T_R}2\left(C-\overline{C}\right)C\frac{\lambda}{\alpha+\lambda}var(\rho). \end{aligned}$$

After the algebra:

Proposition 4 *The price dispersion measured by the standard deviation of equilibrium price distribution is*

$$\sigma_P(t) = \begin{cases} \sqrt{\left[1-e^{-\lambda t}\right]\frac{\alpha}{\alpha+\lambda}\frac{\widetilde{\kappa}_2^2 var(\rho)}{(r+\alpha+\lambda/2)^2} + e^{-\lambda t}\frac{\widetilde{\kappa}_1^2 var(\rho)}{2(r+\lambda/2)^2}\left(C^2+\overline{C}^2-\frac{2C\overline{C}\lambda}{\alpha+\lambda}\right)} & \text{for } t < T_R \\ \sqrt{\left[1-e^{-\lambda(t-T_R)}\right]\frac{\alpha}{\alpha+\lambda}\frac{\kappa_2^2 var(\rho)}{(r+\alpha+\lambda/2)^2} + e^{-\lambda(t-T_R)}\frac{\kappa_1^2}{2(r+\lambda/2)^2}var\left[\theta \mid T_R\right]} & \text{for } t \geq T_R, \end{cases}$$

where

$$
\begin{aligned}
var\left[\theta \mid T_R\right] &= \left[1-e^{-\lambda T_R}\right] \frac{2\alpha}{\alpha+\lambda} \overline{C}^2 var\left(\rho\right) + e^{-\lambda T_R}\left(C^2+\overline{C}^2-2C\overline{C}\frac{\lambda}{\alpha+\lambda}\right) var\left(\rho\right) \\
&+\left(C^2-\overline{C}^2\right) var\left(\rho\right) - \left[1-e^{-(\alpha+\lambda)T_R}\right] 2\left(C-\overline{C}\right)\overline{C}\frac{\lambda}{\alpha+\lambda} var\left(\rho\right) \\
&-e^{-(\alpha+\lambda)T_R} 2\left(C-\overline{C}\right) C\frac{\lambda}{\alpha+\lambda} var\left(\rho\right).
\end{aligned}
$$

One advantage of my model relative to $\{0,1\}$ models of Hugonnier et al. (2014) and Shen et al. (2015) is the following. In $\{0,1\}$ models, the standard deviation of price is not available in closed form, but the difference between the maximum and the minimum price. From an econometric point of view, one would like a measure that takes into account the distributional effect, i.e. trades that are more likely to happen should have higher weight than trades that are less likely, in the calculation of price dispersion. My model allows to calculate the impact of this distributional effect.

Both before and after the recovery, the price dispersion is a weighted average of two terms: first, a dispersion measure of a hypothetical steady-state equilibrium in which investors follow their current trading behavior, and second, the price dispersion at the beginning of the crisis and the recovery period, respectively. Focusing on the first terms, these formulae imply that 3 factors create dispersion in transaction prices in equilibrium. First factor is the distortion on the extensive margin created by the search frictions. Indeed, if there were no search frictions, all investors' marginal valuations would be the same, implying a unique price for all transactions. The second factor is the sensitivity of the price to investor's current intrinsic type, i.e., a price impact which again stems from search frictions. And, the last factor is the heterogeneity in investors' intrinsic types.

The comparative statics analysis of this formula yields interesting empirical implications: To begin with, illiquidity of a market is an important determinant of price dispersion, as stated by other search-theoretic models such as Hugonnier et al. (2014), and it vanishes as the market becomes perfectly liquid (as $\lambda \to \infty$). Gavazza (2011b) empirically finds that price dispersion is higher for illiquid markets. Around $\lambda \simeq \infty$, an increase in λ affects the price dispersion in two different ways: First, since investors find each other faster, the level of misallocation decreases. Because of the resulting decline in the heterogeneity of effective types, trades tend to happen in a less dispersed range of prices. The stronger effect is that a higher λ reduces the sensitivity of prices to current effective types of investors, and in turn decreases the price dispersion.

Secondly, extreme market conditions associated with high volatility of asset payoffs and investor incomes (κ_2) increase the price dispersion. This is consistent with the empirical evidence presented by Friewald et al. (2012) for US corporate bond market and Afonso and Lagos (2012) for the federal funds market that the price dispersion is higher during crises compared to normal market conditions. Finally, investor heterogeneity ($var\left(\rho\right)$) is positively related with price dispersion.

4.2 Trading activity

Instantaneous aggregate trading volume is given by

$$\mathbb{V}\left(t\right) = \frac{1}{2} 2\lambda \mathbb{E}\left[\left|\frac{\theta' - \theta}{2}\right| |t\right]$$

after the recovery. Before the recovery, the same formula applies with replacing θ by $\overline{\theta}$. Thus, it requires using the first absolute moment of individual trade sizes. However, my characterization of equilibrium distribution only allows for usual moments. Consequenlty, I will calculate an approximation, more precisely a sharp upper bound, for the actual trade volume. In the probability theory literature, some sharper bounds for first absolute moments have recently been developed than usual Hölder-Lyapunov inequalities could provide. Using the fact that θ and θ' are distributed i.i.d. due to random matching, Theorem 6 of Ushakov (2011) yields

$$\mathbb{E}\left[\left|\theta' - \theta\right| |t\right] \approx \frac{4}{\pi}\sqrt{var\left[\theta|t\right]}.$$

Therefore,

$$\mathbb{V}\left(t\right) \approx \lambda\frac{2}{\pi}\sqrt{var\left[\theta|t\right]}.$$

Proposition 5 *A sharp upper bound for the instantaneous trading volume in equilibrium is:*

$$\mathbb{V}\left(t\right) \approx \begin{cases} \lambda\frac{2}{\pi}\sqrt{\left[1-e^{-\lambda t}\right]\frac{2\alpha}{\alpha+\lambda}\overline{C}^2 var\left(\rho\right) + e^{-\lambda t}\left(C^2+\overline{C}^2-\frac{2C\overline{C}\lambda}{\alpha+\lambda}\right)var\left(\rho\right)} & \text{for } t < T_R \\ \lambda\frac{2}{\pi}\sqrt{\left[1-e^{-\lambda(t-T_R)}\right]\frac{2\alpha}{\alpha+\lambda}C^2 var\left(\rho\right) + e^{-\lambda(t-T_R)}var\left[\theta \mid T_R\right]} & \text{for } t \geq T_R, \end{cases}$$

where

$$\begin{aligned} var\left[\theta \mid T_R\right] = & \left[1-e^{-\lambda T_R}\right]\frac{2\alpha}{\alpha+\lambda}\overline{C}^2 var\left(\rho\right) + e^{-\lambda T_R}\left(C^2+\overline{C}^2-2C\overline{C}\frac{\lambda}{\alpha+\lambda}\right)var\left(\rho\right) \\ & +\left(C^2-\overline{C}^2\right)var\left(\rho\right) - \left[1-e^{-(\alpha+\lambda)T_R}\right]2\left(C-\overline{C}\right)\overline{C}\frac{\lambda}{\alpha+\lambda}var\left(\rho\right) \\ & -e^{-(\alpha+\lambda)T_R}2\left(C-\overline{C}\right)C\frac{\lambda}{\alpha+\lambda}var\left(\rho\right). \end{aligned}$$

Proposition 5 is important because of its implications about the trading activity in the short term and long term after an aggregate uncertainty shock. In the long term, a flight from this market occurs if

$$\overline{C} < C,$$

i.e., if the steady-state trading volume is bigger than the long term trading volume during the crisis period. In the short term, a decline in the trading activity in this market is observed if

$$C^2+\overline{C}^2-2C\overline{C}\frac{\lambda}{\alpha+\lambda} < \frac{2\alpha}{\alpha+\lambda}C^2.$$

Figure 1 shows the changes in trading activity for various levels of $\overline{C}$.

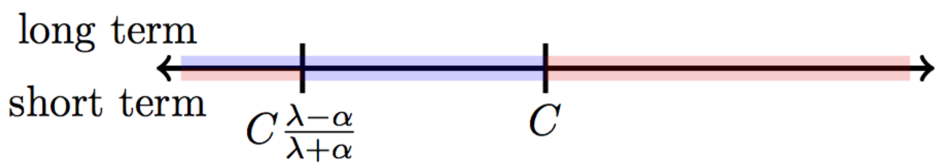

Figure 1. Changes in the trading activity after the aggregate shock at $t=0$ for various $\overline{C}$, where red and blue represent an increase and a decrease, respectively

As discussed earlier, C and $\overline{C}$ are measures of how aggressively investors trade, respectively, during normal times and during crisis. Accordingly, Figure 1 shows that when $\overline{C} > C$, during the crisis, a higher trading activity is observed, and vice-versa. However, during the onset of the crisis, a higher trading activity might be observed if $\overline{C}$ is low enough. This type of "heating-up" in the trading activity is observed in some real-life financial markets and are mainly caused by fire sales. This is also the case in my model. When $\overline{C}$ is low enough, investors suddenly become very cautious. They start to substantially dislike the extreme positions they had before the aggregate shock. This market-wide high tendency towards cautious positions leads to fire sales. Many trades occur at extreme prices.

The fundamental determinant of the relative positioning of C and $\overline{C}$ is the relative positioning of the change, created by the aggregate shock, in the volatility of asset payoff, $\frac{\overline{\sigma}_D}{\sigma_D}$, and in the volatility of background risk, $\frac{\overline{\sigma}_\eta}{\sigma_\eta}$. Define

$$\underline{Y} \equiv \frac{\left(r+\alpha+\frac{\lambda}{2}\right)\left(r+\frac{\lambda}{2}+R\right)\frac{\overline{\sigma}_\eta}{\sigma_\eta}}{2\left(r+\frac{\lambda}{2}\right)\left(r+\alpha+\frac{\lambda}{2}+R\right)} - \frac{\sqrt{\left(\left(r+\alpha+\frac{\lambda}{2}\right)\left(r+\frac{\lambda}{2}+R\right)\frac{\overline{\sigma}_\eta}{\sigma_\eta}\right)^2 - 4\left(r+\frac{\lambda}{2}\right)\left(r+\alpha+\frac{\lambda}{2}+R\right)R\alpha}}{2\left(r+\frac{\lambda}{2}\right)\left(r+\alpha+\frac{\lambda}{2}+R\right)}$$

$$\overline{Y} \equiv \frac{\left(r+\alpha+\frac{\lambda}{2}\right)\left(r+\frac{\lambda}{2}+R\right)\frac{\overline{\sigma}_\eta}{\sigma_\eta}}{2\left(r+\frac{\lambda}{2}\right)\left(r+\alpha+\frac{\lambda}{2}+R\right)} + \frac{\sqrt{\left(\left(r+\alpha+\frac{\lambda}{2}\right)\left(r+\frac{\lambda}{2}+R\right)\frac{\overline{\sigma}_\eta}{\sigma_\eta}\right)^2 - 4\left(r+\frac{\lambda}{2}\right)\left(r+\alpha+\frac{\lambda}{2}+R\right)R\alpha}}{2\left(r+\frac{\lambda}{2}\right)\left(r+\alpha+\frac{\lambda}{2}+R\right)}$$

$$\underline{y} \equiv \frac{\left(r+\alpha+\frac{\lambda}{2}\right)\left(r+\frac{\lambda}{2}+R\right)(\lambda+\alpha)\frac{\overline{\sigma}_\eta}{\sigma_\eta}}{2\left(r+\frac{\lambda}{2}\right)\left(r+\alpha+\frac{\lambda}{2}+R\right)(\lambda-\alpha)} - \frac{\sqrt{\left(\frac{2(r+\alpha)+\lambda}{2}\left(r+\frac{\lambda}{2}+R\right)(\lambda+\alpha)\frac{\overline{\sigma}_\eta}{\sigma_\eta}\right)^2 + 4\frac{2r+\lambda}{2}\frac{2(r+\alpha+R)+\lambda}{2}(\lambda-\alpha)(2r+\alpha+2R)R\alpha}}{2\left(r+\frac{\lambda}{2}\right)\left(r+\alpha+\frac{\lambda}{2}+R\right)(\lambda-\alpha)}$$

$$\overline{y} \equiv \frac{\left(r+\alpha+\frac{\lambda}{2}\right)\left(r+\frac{\lambda}{2}+R\right)(\lambda+\alpha)\frac{\overline{\sigma}_\eta}{\sigma_\eta}}{2\left(r+\frac{\lambda}{2}\right)\left(r+\alpha+\frac{\lambda}{2}+R\right)(\lambda-\alpha)} + \frac{\sqrt{\left(\frac{2(r+\alpha)+\lambda}{2}\left(r+\frac{\lambda}{2}+R\right)(\lambda+\alpha)\frac{\overline{\sigma}_\eta}{\sigma_\eta}\right)^2 + 4\frac{2r+\lambda}{2}\frac{2(r+\alpha+R)+\lambda}{2}(\lambda-\alpha)(2r+\alpha+2R)R\alpha}}{2\left(r+\frac{\lambda}{2}\right)\left(r+\alpha+\frac{\lambda}{2}+R\right)(\lambda-\alpha)}.$$

Figure 2 shows the changes in trading activity for various levels of $\frac{\overline{\sigma}_D}{\sigma_D}$.

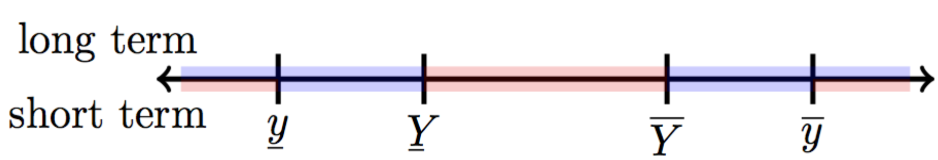

Figure 2. Changes in the trading activity after the aggregate shock at $t = 0$ for various $\frac{\overline{\sigma}_D}{\sigma_D}$, where red and blue represent an increase and a decrease, respectively

The first and second part are actually irrelevant because $\frac{\overline{\sigma}_D}{\sigma_D} < \underline{Y}$ is impossible when $\frac{\overline{\sigma}_\eta}{\sigma_\eta} > 1$. Let me first focus on the long term. When the increase in the volatility of asset payoff is higher than a certain threshold ($\frac{\overline{\sigma}_D}{\sigma_D} > \overline{Y}$), a flight from this market is observed. This is consistent with the "flight-to-quality" events observed in real-life financial markets during crises. When overall uncertainty in the markets is high, investors trade mostly in relatively safe markets. Similarly, when $\frac{\overline{\sigma}_D}{\sigma_D}$ is low, there is flight-to-quality to this market.

When we focus on the short term, we realize that a flight-to-quality to this market always starts with a heating-up of trading activity in short term. However, flight from this market might start with a dry-up or heating-up depending on the level of increase in asset payoff uncertainty relative to the increase in background risk uncertainty. Very severe flights from the market actually starts with a heating-up where fire sales occur in which investors quickly transition to more cautious positions in the asset.

Table 1: Parameter values

Parameter		Value
Discount rate	r	0.05
Risk aversion	γ	0.01
Expected asset payoff	m_D	6.88
Vol. of asset payoff	σ_D	0.25
Vol. of background risk	σ_η	5967
Vol. of asset payoff (crisis)	$\overline{\sigma}_D$	1.64
Vol. of background risk (crisis)	$\overline{\sigma}_\eta$	13143
Asset supply	A	60160
Search efficiency	λ	37
Intensity of idiosyncratic shocks	α	8.13
Intensity of recovery event	R	2.71
Number of intrinsic types	J	30
Intrinsic types	ρ_j	$-1 + 2\beta^{-1}_{(8,8)}\left([j-1]/J\right)$

$\beta^{-1}_{(\alpha,\beta)}(x)$ refers to the inverse cumulative function of a beta distribution with an alpha parameter of α and a beta parameter of β.

4.3 Application to the corporate bond market

In this section, I present a numerical example of my model of crisis to capture the dynamics of price dispersion and trading activity during and after the onset of the subprime crisis in the secondary market for corporate bonds, a typical decentralized asset market. Table 1 shows the parameter values chosen for the calibration.

In the calibration, 1 period is thought of as 1 year. Since the preference structure of my model is same as that of Duffie et al. (2007), I follow them in setting the discount rate to 5% and the risk aversion parameter to 0.01. I normalize the asset supply to $A = 60,170$ so that the average price in the steady-state equilibrium is $\mathbb{E}[P] = 100$.[2] The expected asset payoff, $m_D = 6.88$, and the volatility of asset payoff, $\sigma_D = 0.25$, are chosen to match the average yield spread of $m_D/\mathbb{E}_\phi[P] - r = 1.88\%$ and the standard deviation of yield spread of $\sigma_D/\mathbb{E}_\phi[P] = 0.25\%$ during the period before the subprime crisis, i.e., between February 2006 and June 2007 (Friewald, Jankowitsch & Subrahmanyam, 2012). I do not make adjustments to m_D during the crisis as the coupon rates of corporate bonds before and during the crisis are roughly the same as Friewald et al. (2012) report. I calculate the average price during crisis by targeting the average yield spread of $m_D/\mathbb{E}_\phi\left[\overline{P}\right] - r = 5\%$. To choose the volatility of asset payoff $\overline{\sigma}_D$ during the crisis, I target the standard deviation of yield spread of $\overline{\sigma}_D/\mathbb{E}_\phi\left[\overline{P}\right] = 2.38\%$ during the subprime crisis, i.e., between July 2007 and December 2008 (Friewald et al., 2012).

Friewald et al. (2012) find that the trading interval is 3.38 day before the crisis and 3.37 day during the crisis. In my model, the search efficiency parameter, $\lambda = 37$, implies a trading interval of $\frac{250}{2\lambda} = 3.38$ day. In the calibration of the idiosyncratic shocks in my numerical example, I target the average length of intermediation chains and the price dispersion observed

[2] When I scale up (down) m_D and σ_D, and scale down (up) A by the same constant, all equilibrium objects I calculate for my numerical exercise stay the same. That is, if $\{q, P, \Phi(\rho, a)\}$ is an equilibrium when the asset supply is A, the expected asset payoff is m_D and the asset payoff volatility is σ_D, then, for any $k > 0$, $\left\{\frac{q}{k}, kP, \frac{1}{k}\Phi(\rho, ka)\right\}$ is an equilibrium when the asset supply is $\frac{A}{k}$, the expected asset payoff is km_D, and the asset payoff volatility is $k\sigma_D$.

in the corporate bond market. The intermediation chain length can be written as

$$\mathbb{CL} = \frac{Total\ Trade\ Vol.}{Fundamental\ Vol.}$$

where total trade volume is calculated as in the previous chapter and fundamental volume is the part of total trade volume with trades caused by idiosyncratic shocks, i.e.,

$$Fundamental\ Vol. = \frac{\alpha}{2} C \mathbb{E}\left[\left|\rho' - \rho\right|\right].$$

Using Theorem 6 of Ushakov (2011), a proxy for fundamental volume is

$$Fundamental\ Vol. \approx \frac{\alpha}{2} C \frac{4}{\pi} \sqrt{var\left[\rho\right]}.$$

Hence, we arrive at the following proxy for the intermediation chain length

$$\mathbb{CL} \approx \frac{\sqrt{2}\lambda}{\sqrt{\alpha\left(\alpha+\lambda\right)}}.$$

According to this formula, the chosen intensity of idiosyncratic shocks, $\alpha = 8.13$, leads to an average intermediation chain length of 2.73 which is the one reported by Shen et al (2016) for the period between July 2002 and December 2012. Proposition 4 and 5 imply that scaling σ_η and σ_ρ up and down, respectively, by the same factor has no impact on price dispersion and trading activity. Hence, I normalize σ_ρ to 1/3. To calibrate σ_η, I target the price dispersion of 39.75 *bp* which is the observed one in the corporate bond market during the period before the subprime crisis (Friewald et al., 2012). The average VIX index during the period before crisis was 13.21, while it was 29.1 during the crisis. Thus, I set $\overline{\sigma}_\eta = \frac{29.1}{13.21}\sigma_\eta$. Since the CAPM beta of iShares Barclays Aggregate Bond Fund, one of the most established corporate bond indices, is -0.03, I assume that $\overline{\rho} = 0$, i.e., there is no correlation, on aggregate, between the payoff of corporate bonds and the investors' stochastic income. In the calibration of the distribution, F, of intrinsic types, I assume that it has a generalized beta distribution on $[-1, 1]$. Hence, the mean and standard deviation I chose earlier fully specifies it.

Finally, I assume that $R = \alpha/3$, i.e., investors expect to receive three idiosyncratic shock, on average, before the recovery from the aggregate uncertainty shock. Figure 3 shows the effect of this aggregate uncertainty shock on price dispersion and trading activity.

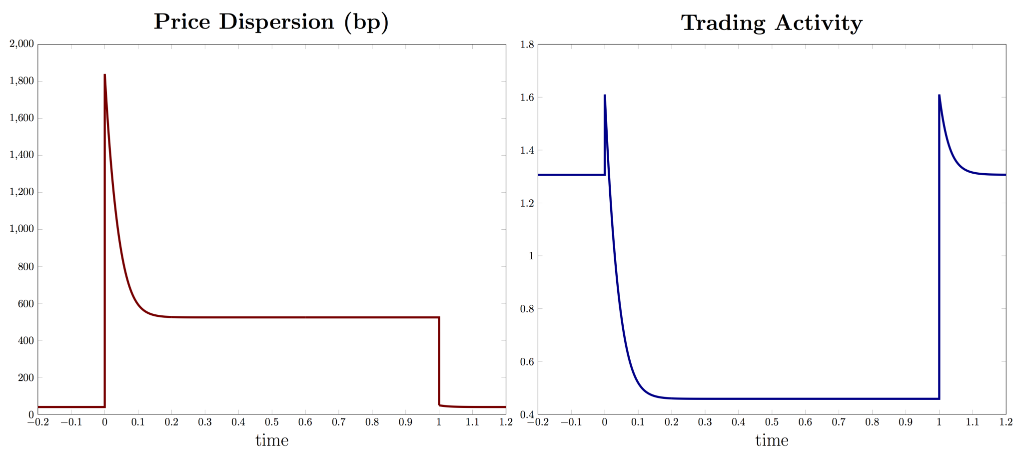

Figure 3. (a) Time-series of the price dispersion (bp) (b) Time-series of the proxy for asset turnover

Since the calibrated values imply that $\overline{C} < C$, we observe a flight-to-quality from the corporate bond market during the aggregate uncertainty shock in the long term. On the other hand, as $\overline{C} < C\frac{\lambda-\alpha}{\lambda+\alpha}$ also holds, a heating-up during the onset of the crisis is observed. Because the investors' new trading regime after $t = 0$ in the corporate bond market is much more cautious than their behavior during the normal times, they have a strong inclination to holding conservative positions after the aggregate uncertainty shock. This leads to plenty of fire sales initially. This, in turn, creates a spike in price dispersion and a "heating-up" in the trading activity. The short heating-up period is followed by a long "dry-up" in the trading activity, which lasts until the recovery at $t = 1$.

5 Conclusion

Assets which became a center of attention with the subprime crisis, like CDOs, CDSs, and repos, are traded in decentralized markets. As a result, analyzing the effect of crises in this type of markets became a center of attention as well. In this paper, I construct a dynamic model of crises in a decentralized asset market that operates via search and bargaining. The crisis is modeled as a one-time aggregate shock to uncertainty with a random recovery. I analyze the effect of this crisis on price dispersion and trading activity. I show that both volatility of asset payoff and volatility of background risk contribute to higher level of price dispersion during the crisis. Trading activity might be higher or lower depending on the increase in the volatility of background risk relative to the increase in the volatility of asset payoff, consistent with the "flight-to-quality" observations during extreme episodes. A flight to the asset market always starts with a "heating-up"in trading activity but a flight from the market might start with a dry-up or heating-up during the onset of the crisis. If the relative increase in the volatility of asset payoff is too high, a period of fire sales is triggered leading to a short heating-up before the complete dry-up of the trading activity. I calibrate the model according to the U.S. corporate bond market data and show that it captures the observations during the subprime crisis.

References

[1] Afonso, G., & Lagos, R. (2012). An empirical study of trade dynamics in the fed funds market. *Federal Reserve Bank of New York Staff Report.*

[2] Bracewell, R. N. (2000). *The Fourier transform and its applications.* New York, NY: McGraw Hill.

[3] Duffie, D., Gârleanu, N., & Pedersen, L. H. (2005). Over-the-counter markets. *Econometrica, 73*, 1815–1847.

[4] Duffie, D., Gârleanu, N., & Pedersen, L. H. (2007). Valuation in over-the-counter markets. *Review of Financial Studies, 20*, 1865–1900.

[5] Friewald, N., Jankowitsch, R., & Subrahmanyam, M. G. (2012). Illiquidity or credit deterioration: A study of liquidity in the US corporate bond market during financial crises. *Journal of Financial Economics, 105*(1), 18-36.

[6] Gârleanu, N. (2009). Portfolio choice and pricing in illiquid markets. *Journal of Economic Theory, 144*(2), 532–564.

[7] Hugonnier, J., Lester, B. & Weill, P.-O. (2014). Heterogeneity in decentralized asset markets. *Mimeo.*

[8] Lagos, R., Rocheteau, G., & Weill, P.O. (2011). Crises and liquidity in over-the-counter markets. *Journal of Economic Theory, 146*(6), 2169–2205.

[9] Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory.* Oxford, UK: Oxford University Press.

[10] Nash, J. F. (1950). The bargaining problem. *Econometrica, 18*(2), 155–162.

[11] Protter, P. (2004). *Stochastic integration and differential equations.* New York, NY: Springer.

[12] Shen, J., Wei, B., & Yan, H. (2015). Financial intermediation chains in an OTC market. *Mimeo.*

[13] Ushakov, N. G. (2011). Some inequalities for absolute moments. *Statistics and Probability Letters, 81*, 2011-2015.

[14] Vayanos, D., & Weill, P.-O. (2008). A search-based theory of the on-the-run phenomenon. *Journal of Finance, 63*, 1361–1398.

[15] Weill, P.-O. (2007). Leaning against the wind. *Review of Economic Studies, 74*(4), 1329–1354.

CHAPTER 3
Endogenous Liquidity and Cross-section of Returns in Dynamic Bargaining Markets

1 Introduction

Analysis of market liquidity (i.e., the ease of buying and selling) as a pricing factor has become a focus of attention in financial research as the recent developments in financial markets and the 2007-2008 subprime crisis outlined the importance of market liquidity for healthy functioning of financial markets and, in turn, of the whole economy. One important question yet to be answered is how market liquidity interact with asset fundamentals. In this paper, I aim at answering this question from the point of view of a model with endogenous liquidity. More precisely, I look, in the theoretical environment provided by the model, at how differences in asset quality lead to differences in market liquidity in the cross-section.

The main motivation comes from the interesting results that empirical cross-sectional analysis of the relation between liquidity and uncertainty yields. Barinov (2014) studies monthly returns on individual stocks traded on NASDAQ over the period 1964-2006. He finds that the turnover (trading volume over shares outstanding) of an asset is positively related to its firm-specific uncertainty.[1] Li and Wu (2006) study daily returns on individual stocks listed in the Dow Jones 30 index, over the period 1988-2001. They find that the trading volume and bid-ask spread of an asset is positively related to its return volatility.[2] However, if turnover and trading volume measure liquidity while bid-ask spread measures illiquidity, the findings of Barinov (2014) and Li and Wu (2006) are puzzling.

Furthermore, controlling for payoff uncertainty, the empirical analysis of liquid/illiquid

[1] See also Comiskey, Walkling, and Weeks (1987) and Karpoff (1987).

[2] See also Schwert (1989), Gallant, Rossi, and Tauchen (1992), and Lang, Litzenberger, and Madrigal (1992).

asset pairs reveals the existence of a return differential, a *liquidity premium*, between assets. The time variation in liquidity premia is delineated by the term "flight-to-liquidity," meaning that liquidity premia are higher during extreme market episodes. For instance, Kamara (1994) studies the yield spread between T-notes and T-bills with matched maturities, which is essentially a liquidity premium stemming from the lower liquidity of notes. He finds that the liquidity premium is positively correlated with interest-rate volatility.[3] Longstaff (2004b) studies the yield spread between the securities of Refcorp and Treasury securities, which is again a form of liquidity premium caused by the lower liquidity of the Refcorp securities. He finds that the liquidity premium is negatively correlated with consumer confidence index.

In order to rationalize those empirical findings, I propose a dynamic bargaining market model in which assets differ in (i) their supply, represented by the quantity of tradeable shares and (ii) their quality, represented by the payoff volatility. My model generates endogenous cross-sectional liquidity differentials consistent with the empirical regularities mentioned above. In addition, I show that times of high volatility are associated with a flight-to-liquidity.

The modeling strategy follows closely Weill (2008). Trade is subject to search and bargaining frictions. Investors optimally choose a portfolio of search efforts. Marginal value of increasing the search effort allocated to a particular asset is affected by the ease of finding a counterparty who holds that asset. Investors recognize this fact. Under natural technical conditions, this fact leads to a trade-off between liquidity and price of assets. That is, controlling for risk, an asset that is easier to find is sold at a higher price.

The first contribution of this paper is theoretical: It extends the risk-neutral search-based pricing model of Weill (2008) so as to treat the implications of search frictions for risky asset pricing. I show how risk aversion can be approximated in a risk-neutral multi-asset setting by means of a quasilinear utility over asset position that accounts for the

[3] See also Amihud and Mendelson (1991) and Strebulaev (2003).

utility reductions stemming from suboptimal hedging. Then, I study how risk and supply differences affect prices and liquidity of assets in a steady-state equilibrium in which investors face idiosyncratic risk. I show the effect of risk aversion on cross-section of asset prices in a setting with search, above and beyond the usual implications of search-theoretical models in finance. The analysis in this paper could not have been conducted in Weill (2008) in which the cross-sectional variation in asset returns is exclusively due to liquidity differences and risk is not priced.

The second contribution of this paper is to classify the measures of liquidity. I show that cross-sectional liquidity and risk are negatively related in the case of buyers, while they are positively related for sellers. Combining these results with the results of Weill (2008) regarding supply-liquidity relationship, I am able to propose the following measures as proxies for liquidity. For buyers, I suggest that bid-ask spread is a proxy for illiquidity while market capitalization is a proxy for liquidity. In the case of sellers, trading volume and turnover are proxies for liquidity. Consequently, higher risk implies lower liquidity for buyers and hence higher bid-ask spread and lower market capitalization. On the other hand, higher risk implies higher liquidity for sellers and hence higher trading volume and turnover. These results are consistent with empirical regularities mentioned before.

The last section of the paper addresses time variation in liquidity. First, I demonstrate how my model may lead to flight-to-liquidity. In demonstration of flight-to-liquidity, I characterize an equilibrium in which the cross-sectional variation in asset returns is exclusively due to liquidity differences as is the case in Weill (2008). I show that liquidity premia increase with common payoff volatility of assets suggestive of flight-to-liquidity.

Early applications of search-theoretical tools can be found in labor economics, spurred by the "coconuts" model of Diamond (1982). Then, they are also used to answer important questions in monetary economics.[4] As for search-theoretical models of finance, Duffie, Gâr-

[4]See, for instance, Trejos and Wright (1995).

leanu, and Pedersen (2005)[5] offer an over-the-counter (OTC) market model with risk-neutral investors in which only one asset is traded. Miao (2006) generalizes this model to study the co-existence of centralized and OTC markets. Vayanos and Wang (2007) and Weill (2008) have provided important generalizations to an environment with many assets. Weill (2007) shows how dealers provide liquidity during a crisis in an extension of Duffie et al. (2005). Longstaff (2004a) and Gârleanu (2006) propose models of infrequent trading by replacing the bilateral bargaining assumption with an infrequent access to a centralized market. Krainer and LeRoy (2002) offer a different search-theoretical framework to study the housing market. Finally, my paper is also related to asset pricing models with exogenously specified trading costs.[6] More precisely, I complement this literature with a model of endogenous trading costs in the context of search frictions.

The remainder of the paper is organized as follows. Section 2 describes the setup, Section 3 defines, calculates, and analyzes an equilibrium where buyers search for all assets. Section 4 discusses the results and study a flight-to-liquidity event. Lastly, Section 5 concludes. The Appendix provides the microfoundations for the risk-adjusted utility employed in my analysis.

2 Environment

This section presents the model environment, in which investors encounter two non-Walrasian features in the dynamic bargaining market where they trade. First, locating a counterparty takes time and effort. Investors allocate "trading specialists" to search for asset-specific counterparties. Second, when two trading specialists meet, they negotiate over the price on behalf of the investors they work for. The model setup is adapted from Weill (2008). The

[5] See also Duffie, Gârleanu, and Pedersen (2002) and Duffie, Gârleanu, and Pedersen (2007).

[6] See Amihud and Mendelson (1986), Constantinides (1986), Vayanos (1998), Huang (2003), Vayanos (2004), and Acharya and Pedersen (2005).

novel part of my setup is that the preferences are generalized in a way to account for the asset payoff risk so that not only liquidity is a pricing factor but the risk is also priced in the equilibrium of this model.

2.1 Preferences

Time is continuous and runs forever. I fix a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\{\mathcal{F}_t, t \geq 0\}$ of sub-σ-algebras satisfying the usual conditions (see Protter, 2004). The filtration represents the resolution of information commonly available to investors over time. Multiple assets $k \in \{1, 2, ..., K\}$ are traded. Fixed supply of asset k is given by $s_k \in (0, 1)$. Suppose δ and σ_k are positive constants, and B_t is a standard Brownian motion. The cumulative dividend process

$$dD_{kt} = \delta dt + \sigma_k dB_t \tag{1}$$

describes the cash flow paid by asset k.

The economy is populated by a continuum of infinitely lived investors with total measure normalized to 1. Investors' time preferences are determined by a constant discount rate $r > 0$. An investor is permitted to hold either zero or one share of some asset, and can choose which asset to hold. Investors are effectively risk-averse. In order to stay away from the burden of derivations with minimal intuition, however, I assume investors are risk-neutral but receive a risk-adjusted utility flow from holding a position in an asset.

At any point in time, an investor has intrinsically either a high valuation or a low valuation for holding assets. When his intrinsic-type is high and holds asset $k \in \{1, 2, ..., K\}$, he enjoys the utility flow $\delta - A\sigma_k^2$ where A is a positive parameter. With a low valuation, he enjoys a utility flow $\delta - x\sigma_k - A\sigma_k^2$, for some parameter $x > 0$.[7] Any investor's intrinsic type switches

[7] These utility flows are interpreted in terms of risk aversion. Since the parameter δ is an expected rather than actual dividend flow, this cash flow needs to be adjusted for risk. The term $A\sigma_k^2$ represents a cost of risk bearing. The term $x\sigma_k$ represents additional holding cost for low-type investors. In the Appendix, I provide microfoundations for these utilities. I assume that investors have CARA preferences over the

from low to high with intensity γ_u, and switches back with intensity γ_d. For any pair of investors, their intrinsic type processes are assumed to be independent. In addition to the utility flow provided by the assets, an investor cares for the consumption of a nonstorable *numéraire* good, with a marginal utility of 1. Investors are endowed with a technology that instantly produces the numéraire at unit marginal cost so that they are able to make side payments in the numéraire.

I let $s \equiv (s_1, s_2, ..., s_K)$ denote the vector of asset supplies. I also impose the condition that

$$\sum_{k=1}^{K} s_k \equiv S < \frac{\gamma_u}{\gamma_u + \gamma_d}, \tag{2}$$

which implies that the total number of tradeable shares S of all assets is less than the measure of high-type investors in the steady-state. As investors are allowed to hold 0 or 1 unit of some asset, Assumption (2) implies that the so-called marginal investor of a hypothetical frictionless market would be of high type. Hence, in this frictionless benchmark, asset k would have the equilibrium price of $\frac{\delta - A\sigma_k^2}{r}$ since, at any point in time, assets would be allocated to those who value them the most.

2.2 Trade

At any point in time, investors differ from each other in two characteristics: an intrinsic type (high h or low l) and whether he owns an asset or not (owner of asset k ok or non-owner n). Therefore, the full set of investor types is

$$I = \{hn, ln, ho1, ..., hoK, lo1, ..., loK\}. \tag{3}$$

numéraire good, and that they can invest in a riskless asset with return r and in two different risky assets with dividend flow described in Equation (1). Moreover, investors receive a stochastic income flow whose correlation with the dividend flow can be positive (low-type) or zero (high-type). These assumptions give rise to the risk-adjusted utility flows for low-type and high-type investors, with the parameters A and x being functions of the investors' risk aversion and the background risk correlation.

As implied by Assumption (2), high-type non-owners (hn) will be potential buyers, and low-type owners of asset k (lok) will be potential sellers of asset k in equilibrium. For each $i \in I$, I let μ_i denote the fraction of investors of type i. Every investor is endowed with a measure $\overline{v}$ of "trading specialists" who search for and negotiate with trading counterparties on behalf of investors.

A trading specialist is characterized by a two-dimensional vector $(i, j) \in I^2$, meaning that he works for an investor of type i and is allocated by the investors to searching specialists who work for investors of type j. Therefore, the total measure of specialists of type (i, j) is $\mu_i v_{ij}$. Any investor's choice of trading specialist allocation is constrained by his total endowment of trading speciaists such that $\sum_{j \in I} v_{ij} \leq \overline{v}$ for all $i \in I$. A given specialist finds a counterparty with an intensity $z > 0$, reflecting the overall liquidity in the dynamic bargaining markets. Contacts are also pair-wise independent with the investor's intrinsic type processes. Specialists of type (i, j) target only the specialists of type (j, i) as contacts, that could result in a trade, can occur only between those types of specialists. I assume that the counterparty found is randomly selected from the pool of all specialists. Since one can scale $\overline{v}$ and z up and down by the same factor, respectively, I normalize the measure of investors' trading specialist endowment $\overline{v}$ to 1. Thus, for a specialist of type (i, j), conditional on a contact, the random matching assumption implies that the probability that the counterparty is indeed a targeted one is $\mu_i v_{ij}$. As a result, assuming that the law of large numbers applies, specialists of type (i, j) contact specialists of type (j, i) at an (almost sure) instantaneous rate of $\mu_i v_{ij} z \mu_j v_{ji}$ and, since specialists of type (j, i) contact specialists of type (i, j) at the same intantaneous rate, the total rate of such counterparty matching, for $i \neq j$, is

$$\mu_i v_{ij} z \mu_j v_{ji} + \mu_j v_{ji} z \mu_i v_{ij} = 2z \mu_i v_{ij} \mu_j v_{ji} \tag{4}$$

In a discrete-time search-and-matching environment, Duffie and Sun (2007) show that the exact law of large numbers for a continuum of investors indeed applies in the sense that is presented above. Giroux (2005) provides a proof that the cross-sectional distribution of investor types in a natural discrete-time analogue of this type of models indeed converges to its continuous-time counterpart like the one studied here.

3 Equilibrium

This section defines and characterizes a symmetric equilibrium in the sense that potential buyers optimally search for all assets, and analyzes the basic properties of this equilibrium.

3.1 Definition

The equilibrium definition I use is identical to the one defined in Weill (2008). After a brief description of the notation and the anticipated equilibrium trading behavior, first, I will define the investors' value functions, taking as given the equilibrium distribution of investor types. Then, I will write down the conditions that the equilibrium distribution of investor types satisfies. Lastly, I will define the equilibrium.

3.1.1 Individual trades

Let V_i denote the maximum attainable continuation utility of an investor of type i. Reservation value of a high-type non-owner for asset k is defined as $\Delta V_{hk} \equiv (V_{hok} - V_{hn})$. Similarly, reservation value of a low-type owner for asset k is $\Delta V_{lk} \equiv (V_{lok} - V_{ln})$. The total surplus created by a trade between two investors of type hn and type lok, thus, becomes

$$\Delta V_{hk} - \Delta V_{lk} = (V_{hok} - V_{hn}) - (V_{lok} - V_{\mathrm{ln}}). \tag{5}$$

In equilibrium, it will be the case that $\Delta V_{hk} - \Delta V_{lk} > 0$ for all $k \in \{1, 2, ..., K\}$. The surplus of a trade can also be positive between a low-type owner lok and a high-type owner hoj. One of the two might be willing to transfer a specific amount of the numéraire good to the other in order to swap their initial assets. The total surplus of such a swap is $V_{loj} - V_{lok} + V_{hok} - V_{hoj}$.

In my equilibrium definition, I do not allow for swaps. More precisely, I impose the condition that lok investors search only for a direct sale with hn investors, but do not search for swaps. On the sell side of the market, an lok investor allocates all of his trading specialists to search for hn investors, i.e., $v_{lok,hn} = 1$ for all $k \in \{1, 2, ..., K\}$. On the buy side of the market, an hn investor allocates a measure $v_{hn,lok} > 0$ of his trading specialists to search for asset $k \in \{1, ..., K\}$ held by lok investors. Accordingly, when I use the term *trading specialists allocation* in the rest of the paper, I always refer to a trading specialists allocation by hn investors as given in the following definition.

Definition 1 *A trading specialists allocation is some* $v \in \mathbb{R}_+^K$ *with* $\sum_{k=1}^{K} v_k \leq 1$.

In order to make sure that, in the equilibrium, an lok investor does not search for swaps, we need to check if the expected value of searching for a swap is strictly less than the expected value of searching for a direct sale. This condition can be written as

$$2zv_{jk}\mu_{hoj}q(V_{loj} - V_{lok} + V_{hok} - V_{hoj}) < 2zv_k\mu_{hn}q(\Delta V_{hk} - \Delta V_{lk}) \tag{6}$$

for all $(k, j) \in \{1, ..., K\}^2$, where v_{jk} is the measure of trading specialists that hoj investors allocate to search for a swap with lok investors. Dividing both sides by $2zq$, $v_{jk} \leq 1$ implies that (6) will hold if

$$\mu_{hoj}(V_{loj} - V_{lok} + V_{hok} - V_{hoj}) < v_k\mu_{hn}(\Delta V_{hk} - \Delta V_{lk}) \tag{7}$$

for all $(k, j) \in \{1, ..., K\}^2$. This sufficiency condition will be verified in the proof of Proposition 3. Hence, searching for an direct sale with an hn investor becomes a dominant strategy for lok investors in equilibrium.

The price of every transaction is determined through bilateral bargaining. The lok investor will sell his asset to the hn investor, in exchange for some payment p_k in the numéraire good. The price arises as the solution of a generalized Nash-bargaining game, as follows:

$$p_k = \arg\max_{p} \ (\Delta V_{hk} - p)^{1-q} (p - \Delta V_{lk})^q$$

for all $k \in \{1, ..., K\}$, where $q \in (0, 1)$ is the bargaining power of the lok investor. Solution of the optimization problem above gives that the bargaining process results in the price of asset k:

$$p_k = q\Delta V_{hk} + (1-q)\Delta V_{lk}. \tag{8}$$

3.1.2 HJB equations

I define the equilibrium continuation utilities of investors recursively. The Hamilton-Jacobi-Bellman (HJB) equation for the continuation utility of a buyer hn is

$$rV_{hn} = \max_{\widetilde{v}_1, \widetilde{v}_2, ..., \widetilde{v}_K} \{\gamma_d(V_{ln} - V_{hn}) + \sum_{k=1}^{K} 2z\widetilde{v}_k \mu_{lok}(V_{hok} - V_{hn} - p_k)\}, \tag{9}$$

subject to

$$\sum_{k=1}^{K} \widetilde{v}_k = 1$$

and

$$\widetilde{v}_k \geq 0,$$

for all $k \in \{1, 2, ..., K\}$. A buyer takes as given the equilibrium measures of investor types but he is able to choose freely how he allocates his trading specialists to different assets.

Following from the Bellman principle of optimality in a stationary equilibrium, the HJB equation (9) equates the flow continuation utility rV_{hn} into the sum of two terms. The first term, $\gamma_d(V_{\mathrm{ln}} - V_{hn})$, is the expected change in the flow utility caused by an idiosyncratic shock. The second term is the expected contribution of searching for alternative assets to the flow utility.

Other investors' continuation utilities solve the following system of HJB equations by taking as given the distribution of investor types and the trading specialists allocation of buyers:

$$rV_{hok} = \delta - A\sigma_k^2 + \gamma_d(V_{lok} - V_{hok}) \tag{10}$$

$$rV_{lok} = \delta - x\sigma_k - A\sigma_k^2 + \gamma_u(V_{hok} - V_{lok}) + 2zv_k\mu_{hn}(V_{ln} - V_{lok} + p_k) \tag{11}$$

$$rV_{ln} = \gamma_u(V_{hn} - V_{ln}) \tag{12}$$

for all $k \in \{1, 2, ..., K\}$.

3.1.3 Stationary distribution of types

I now provide the conditions that the stationary distribution of investor types should satisfy. Since $\mu = (\mu_{hn}, \mu_{hok}, \mu_{lok}, \mu_{\mathrm{ln}})_{1 \le k \le K}$ is a probability mass function, it should satisfy

$$1 = \sum_{k=1}^{K}(\mu_{lok} + \mu_{hok}) + \mu_{hn} + \mu_{ln}. \tag{13}$$

Clearing of the market for all assets requires that

$$s_k = \mu_{lok} + \mu_{hok}, \tag{14}$$

for all $k \in \{1, ..., K\}$. Lastly, for all $i \in I$, μ_i should satisfy the condition for stationarity. For instance, for hn investors, the condition is

$$\gamma_u \mu_{\ln} = \gamma_d \mu_{hn} + \sum_{k=1}^{K} 2zv_k \mu_{hn} \mu_{lok}. \tag{15}$$

The condition basically equates the inflow from ln investors, who receive an idiosyncratic shock, to the sum of the outflow of hn investors due to an idiosncratic shock and due to trade. The stationarity condition for ln investors can be written similarly:

$$\gamma_d \mu_{hn} + \sum_{k=1}^{K} 2zv_k \mu_{hn} \mu_{lok} = \gamma_u \mu_{\ln}.$$

Note that this condition is actually the same as (15).

Equating the inflow and outflow for lok investors yields

$$\gamma_d \mu_{hok} = \gamma_u \mu_{lok} + 2zv_k \mu_{hn} \mu_{lok} \tag{16}$$

for $k \in \{1, ..., K\}$. The condition for hok investors

$$\gamma_u \mu_{lok} + 2zv_k \mu_{hn} \mu_{lok} = \gamma_d \mu_{hok}$$

is the same (16).

Weill (2008) studies (13)–(16) and provides a proof for the following proposition.

Proposition 1 *Given an allocation v of trading specialists, the system (13)-(16) has a unique solution $\mu = (\mu_{hn}, \mu_{hok}, \mu_{lok}, \mu_{ln})_{1 \leq k \leq K} \in [0, 1]^{2K+2}$.*

I have specified all requirements to define a stationary equilibrium. I only have to bring those ingredients together. Formally, a stationary symmetric equilibrium is defined as follows:

Definition 2 *A stationary symmetric equilibrium is*

i) *a collection of prices* $p = (p_1, ..., p_K)$

ii) *a collection of continuation utilities* $V = (V_{hn}, V_{hok}, V_{lok}, V_{ln})_{1 \leq k \leq K}$

iii) *a distribution of types* $\mu = (\mu_{hn}, \mu_{hok}, \mu_{lok}, \mu_{ln})_{1 \leq k \leq K}$

iv) *a trading specialists allocation* $v = (v_1, ..., v_K) \gg 0$

such that

- *Stationarity: Given iv), iii) solves the system (13)-(16).*
- *Optimality: Given i), iii) and iv), ii) and iv) solves the system (9)-(12) of HJB equations. The no-swap condition (7) holds for all* $(k, j) \in \{1, ..., K\}^2$.
- *Nash-bargaining: i) satisfies (8).*

3.2 Characterization

The key property of this model is the clear trade-off it introduces between the liquidity of an asset and the individual transaction surplus that results from trading that asset. Note that our equilibrium concept imposes the condition that all assets are searched, i.e., $v \gg 0$. The first-order condition of the buyer's problem (9), therefore, implies that buyers should be indifferent between searching for all assets since (9) is a linear program:

$$2z\mu_{lok}(V_{hok} - V_{hn} - p_k) = 2z\mu_{loj}(V_{hoj} - V_{hj} - p_j) \tag{17}$$

$$\Longleftrightarrow 2z\mu_{lok}(1 - q)(\Delta V_{hk} - \Delta V_{lk}) = 2z\mu_{loj}(1 - q)(\Delta V_{hj} - \Delta V_{lj}) \tag{18}$$

for all $(k, j) \in \{1, ..., K\}^2$, where (18) follows from substituting (8) into (17). Using this search indifference, one can simplify the HJB equations as follows. First, the search-indifference conditions (18) can be written as

$$W_k = W, \tag{19}$$

for all $k \in \{1, ..., K\}$, for some positive constant W to be determined in equilibrium, where

$$W_k \equiv 2z\mu_{lok}(1-q)(\Delta V_{hk} - \Delta V_{lk}). \tag{20}$$

To show the existence and uniqueness of the equilibrium, I follow the steps in Weill (2008). I start by replacing (19) in Equation (9). Then I combine the HJB equations (9)–(11) and use the pricing equation (8) to arrive at

$$rW_k = 2z\mu_{lok}(1-q)x\sigma_k - (\gamma_d + \gamma_u + 2z\mu_{hn}v_k q)W_k - 2z\mu_{lok}(1-q)W \tag{21}$$

for all $k \in \{1, ..., K\}$. Substituting $\mu_{hok} = s_k - \mu_{lok}$ into Equation (16), I find that

$$2z\mu_{hn}v_k = \frac{\gamma_d s_k}{\mu_{lok}} - (\gamma_d + \gamma_u) \tag{22}$$

Substituting (22) into (21), using (19) and rearranging gives

$$\frac{r + (1-q)(\gamma_d + \gamma_u)}{(1-q)x\sigma_k}\frac{1}{2z\mu_{lok}} + \frac{2z\gamma_d s_k q}{(1-q)x\sigma_k}\frac{1}{(2z\mu_{lok})^2} + \frac{1}{x\sigma_k} = \frac{1}{W} \tag{23}$$

This quadratic equation allows me to write $2z\mu_{lok} = m_k(W)$, for some $W < \min_k\{x\sigma_k\}$, and for some continuous and increasing function $m_k(.)$.

Now, the stationary measure of high-type investors is equal to the stationary probability of being in a state of high intrinsic type

$$\mu_{hn} + \sum_{k=1}^{K} \mu_{hok} = \frac{\gamma_u}{\gamma_d + \gamma_u} \tag{24}$$

Combining (24) with (14) shows

$$\mu_{hn} = \frac{\gamma_u}{\gamma_d + \gamma_u} - S + \sum_{k=1}^{K} \mu_{lok} \tag{25}$$

Substituting (25) into (22) gives

$$v_k\left(\frac{2z\gamma_u}{\gamma_d + \gamma_u} - 2zS + \sum_{k=1}^{K} m_k(W)\right) = \frac{2z\gamma_d s_k}{m_k(W)} - (\gamma_d + \gamma_u) \tag{26}$$

which shows that $\sum_{k=1}^{K} v_k = 1$ only if

$$2z\left(\frac{\gamma_u}{\gamma_d + \gamma_u} - S\right) + \sum_{k=1}^{K} m_k(W) - 2z\gamma_d \sum_{k=1}^{K} \frac{s_k}{m_k(W)} + K(\gamma_d + \gamma_u) = 0 \tag{27}$$

Since $m_k(.)$ is strictly increasing for each k, the left-hand side of (27) is strictly increasing in W. Thus, the equilibrium W is uniquely characterized by (27). Once W is found, the other equilibrium objects are pinned down uniquely: the trading specialists allocation by (26), the distribution of types by (13)–(16), the continuation utilities V by (9)–(11), and the prices p by (8). This implies

Proposition 2 (Uniqueness) *There is at most one symmetric equilibrium.*

In the proof of existence, I again follow Weill (2008) and first analyze the case of identical asset characteristics, for the distribution $\widehat{s} = (S/K, ..., S/K)$ of asset supplies and for the distribution $\widehat{\sigma} = (\widetilde{\sigma}, ..., \widetilde{\sigma})$ of dividend volatilities. I show the existence of a symmetric equilibrium with $v_k = 1/K$, following Duffie et al. (2005). Then, the application of the Implicit Function Theorem (see Taylor & Mann, 1983, chapter 12) to Equation (27) establishes existence in a neighborhood of this equilibrium with identical asset characteristics.

Proposition 3 (Existence with almost-identical assets)

Let $\widehat{s} = (S/K, ..., S/K)$ and $\widehat{\sigma} = (\widetilde{\sigma}, ..., \widetilde{\sigma})$. Then, there is a neighborhood $N \subset \mathbb{R}_+^K$ of $\widehat{s}$ and a neighborhood $M \subset \mathbb{R}_+^K$ of $\widehat{\sigma}$, such that for all $s \in N$ and $\sigma \in M$, there is a symmetric equilibrium.

Proof. I start by guessing that there is a symmetric equilibrium, with $\widehat{\mu}_{lok} = \widehat{\mu}_{lo}/K$ and $\widehat{v}_k = 1/K$, when assets have identical characteristics. The equations that characterize the equilibrium are identical to those of Duffie et al. (2005), after replacing their "λ" with "z/K" here. Their results imply that $V_i > 0$ for all $i \in \{1, 2, ..., I\}$ and $\Delta V_{hk} - \Delta V_{lk}$ for all $k \in \{1, 2, ..., K\}$. Moreover, the no-swap condition (7) is trivially satisfied as assets have identical characteristics. Since the left-hand side of (27) is strictly increasing in W, the Implicit Function Theorem can be applied: This provides a neighborhood $N \subset \mathbb{R}_+^K$ of $\widehat{s}$ and a neighborhood $M \subset \mathbb{R}_+^K$ of $\widehat{\sigma}$, such that, for all $s \in N$ and $\sigma \in M$, there exists a candidate equilibrium $W = h(s, \sigma)$, for some continuous function $h(.,.)$. The other candidate equilibrium objects (V, μ, v) are easily expressed as continuous functions of W and thus as continuous functions of s and σ. The search-indifference conditions (19) are satisfied by construction. All other relevant inequalities hold by continuity. ■

Specifically, the proof establishes that sellers do not have any incentive to search for swaps if assets characteristics are sufficiently homogeneous. This follows from the fact that the net utility of swapping two assets with nearly identical characteristics is close to zero, and turns out to be strictly less than the value of searching for a direct sale. The proofs so far are almost identical to those in Weill (2008). The main difference lies in the holding cost term of low-type investor. The additional dimension of asset characteristics in my model, which is dividend volatilities, creates cross-sectional differences in the holding costs of assets. When $\sigma_k = \sigma$ for all $k \in \{1, 2, ..., K\}$, my model becomes identical to the model of Weill (2008) up to a relabeling of the preference parameters.

Does there always exist an equilibrium in which all assets are traded? Weill (2008) provides a partial answer, in a two-asset economy. Specifically, he shows that if the assets have sufficiently different supplies, there cannot be an equilibrium in which both are traded. Existence in Proposition 3 is proved by studying how the left-hand side of (27) depends on s. When asset characteristics are sufficiently similar, the equation has a solution. Alternatively, when the supply of an asset is sufficiently small relative to supplies of other assets, (27) has no solution.

4 Applications

4.1 Cross-section of returns

In this section, I focus on the implications of the model for asset pricing. I first discuss how risk and liquidity arise endogenously as pricing factors in the model. Then, I discuss the relation of the cross-sectional variation in asset returns with the exogenous vector $s = (s_1, ..., s_K)$ of asset supplies and the exogenous distribution $\sigma = (\sigma_1, ..., \sigma_K)$ of payoff volatilities. The last objective of this section is to classify the liquidity proxies.

4.1.1 Three equations

The pricing equation (8) can be written

$$p_k = \Delta V_{hk} - (1 - q)(\Delta V_{hk} - \Delta V_{lk}) \tag{28}$$

Using the HJB equations (9) and (10), along with (28), I find that

$$rp_k = \delta - A\sigma_k^2 - \gamma_d(\Delta V_{hk} - \Delta V_{lk}) - W - (1-q)r(\Delta V_{hk} - \Delta V_{lk}) \tag{29}$$

This equation decomposes the "flow" price of asset k into five terms. The first, δ, is the expected flow of dividend payments. The second component, $A\sigma_k^2$, is the flow cost of bearing risk. The third component, $\gamma_d(\Delta V_{hk} - \Delta V_{lk})$, is the expected cost of switching to the low-type, and being stuck with the asset due to search frictions. The fourth component, W, is the opportunity cost of buying the asset, i.e., giving up the expected net benefit of continuing search. The last component is the bargaining discount.

It is instructive to compare the price p_k of the asset in this frictional market with its price p_k^∞ in a hypothetical frictionless market. Assumption (2) implies that the so-called marginal investor of a hypothetical frictionless market would be of high type. Hence, in this frictionless benchmark, asset k would have the equilibrium price of $p_k^\infty = \frac{\delta - A\sigma_k^2}{r}$ since, at any point in time, assets would be allocated to those who value them the most. Hence, all discounts in (29) disappear in the frictionless benchmark, except for the cost of fundamental risk-bearing.

Substituting the search-indifference condition (19) into (29), I derive the first important equation

$$p_k = \frac{\delta - A\sigma_k^2}{r} - \frac{W}{r} - (1 + \frac{\gamma_d}{r(1-q)})\frac{W}{2z\mu_{lok}} \tag{30}$$

This equation implies that, controlling for payoff volatility, an asset that is easier to find (has larger μ_{lok}) is sold at a higher price. The effect of σ_k on the price of asset k is indeterminate at this stage of the analysis, since μ_{lok} is endogenous and we do not know the effect of σ_k on μ_{lok}.

Let E, F, and G be positive constants. Note that the indifference condition (23) has the form

$$E\frac{s_k}{\sigma_k}\frac{1}{\mu_{lok}^2}+F\frac{1}{\sigma_k}\frac{1}{\mu_{lok}}+G\frac{1}{\sigma_k}=\frac{1}{W} \tag{31}$$

This is the second important equation and establishes the relation of the measure μ_{lok} of sellers with the asset supply s_k and the payoff volatility $\sigma_{k.}$

The third equation follows from (22), and establishes the relation of the allocation v_k of trading specialists with the measure μ_{lok} of sellers and the fixed asset supply s_k:

$$\frac{\mu_{lok}}{s_k}=\frac{\gamma_d}{\gamma_d+\gamma_u+2z\mu_{hn}v_k} \tag{32}$$

The object $2z\mu_{hn}v_k$ represents the demand side of the market. The larger $2z\mu_{hn}v_k$ is, the higher the search activity for asset k is, and the easier it is to sell this asset. A natural question is if $2z\mu_{hn}v_k$ is an increasing function of μ_{lok}. In other words, is an asset that is easier to sell is also easier to find? Eq. (32) implies that the answer depends on the asset supply s_k, and is therefore unknown at this stage of the analysis.

4.1.2 Liquidity - risk - return relationships

Eq. (31) is of the form

$$H(s_k,\sigma_k,\mu_{lok})=\frac{1}{W} \tag{33}$$

for some function $H(\cdot,\cdot,\cdot)$ that is increasing in s_k and decreasing in σ_k and μ_{lok}. This implies that μ_{lok} is increasing in s_k and decreasing in σ_k. In other words, controlling for payoff volatility, an asset with higher supply is easier to find, is sold at a higher price, and has a lower return $R_k=\delta/p_k$. Similarly, controlling for supply, an asset with higher payoff volatility is harder to find, is sold at a lower price, and has a higher return $R_k=\delta/p_k$.

s_k affects the price through the measure of sellers. An asset with larger s_k has a larger μ_{lok}. By the main pricing equation (30), this implies a higher price.

σ_k affects the price through two channels. First, it has a negative impact on the investors' utility flow because they are risk averse. Secondly, an asset with larger σ_k has a lower μ_{lok}. By the main pricing equation (30), this implies a lower price. Hence, the existence of search frictions amplifies the pricing impact of the fundamental risk.

The discussion above can be summarized in Proposition 4 and 5.

Proposition 4 *Fixing $\widehat{\sigma} = (\widetilde{\sigma}, ..., \widetilde{\sigma})$, in equilibrium, $s_k > s_j$ implies that*

$$\mu_{lok} > \mu_{loj},\ v_k > v_j\ ,\ p_k > p_j\ ,\ R_k < R_j\ ,\ \text{and}\ \Delta V_{hk} - \Delta V_{lk} < \Delta V_{hj} - \Delta V_{lj}.$$

Proof. By (33), $\mu_{lok} > \mu_{loj}$. Since $\frac{\partial \mu_{lok}}{\partial s_k} < 1$, (32) implies $v_k > v_j$. By (30), $\mu_{lok} > \mu_{loj}$ implies $p_k > p_j$ and $R_k < R_j$. By search indifference, $\mu_{lok} > \mu_{loj}$ implies $\Delta V_{hk} - \Delta V_{lk} < \Delta V_{hj} - \Delta V_{lj}$. ■

In words, controlling for payoff volatility, an asset with higher supply is easier to find, easier to sell, has a higher price, a lower return, and a narrower individual trade surplus. When we fix a particular risk for all assets, return differentials arise exclusively due to liquidity differentials, and our model becomes identical to Weill (2008) model. Hence, our Proposition 4 is identical to the proposition 5 in Weill (2008).

Proposition 5 *Fixing $\widehat{s} = (S/K, ..., S/K)$, in equilibrium, $\sigma_k > \sigma_j$ implies that $\mu_{lok} < \mu_{loj}$,*

$$v_k > v_j\ ,\ p_k < p_j\ ,\ R_k > R_j\ ,\ \text{and}\ \Delta V_{hk} - \Delta V_{lk} > \Delta V_{hj} - \Delta V_{lj}.$$

Proof. By (33), $\mu_{lok} < \mu_{loj}$. (32) implies $v_k > v_j$. By (30), $\sigma_k > \sigma_j$ and $\mu_{lok} > \mu_{loj}$ imply $p_k > p_j$ and $R_k < R_j$. By search indifference, $\mu_{lok} > \mu_{loj}$ implies $\Delta V_{hk} - \Delta V_{lk} > \Delta V_{hj} - \Delta V_{lj}$. ■

In words, controlling for supply, an asset with higher payoff volatility is harder to find, easier to sell, has a lower price, a higher return, and a wider individual trade surplus. This model generates a negative relationship between risk and liquidity for buyers because an asset that is riskier is harder to find. The assumption (2) implies high-type investors are on the long side of the market. In symmetric equilibrium, high-type investors without an asset are buyers who are indifferent between searching for any two assets and low-type investors with an asset are sellers. Then, surplus of a trade is an increasing function of risk mainly because holding cost of low-type investors is positively related to risk. Since an asset that is riskier has a larger trade surplus, it should be less liquid so that buyers are indifferent between all assets. This generates the negative relationship between risk and liquidity for buyers. On the other hand, this model generates a positive relationship between risk and liquidity for sellers because an asset that is riskier is easier to sell. Since an asset that is riskier has a larger trade surplus, buyers search for riskier assets with a larger search intensity. Then, this generates the positive relationship between risk and liquidity for sellers.

4.1.3 Risk and liquidity proxies

In this subsection, I analyze natural proxies for liquidity: bid-ask spread, turnover, trading volume and market capitalization. In particular, I study their dependence on the exogenous "liquidity" factors: the quantity of tradeable shares and the payoff volatilities. Then, I relate them to liquidity for buyers, liquidity for sellers or both.

Bid-ask Spread As in Weill (2008), individual trade surpluses may be interpreted as bid-ask spreads, in the following sense. Suppose a "monopolistic" marketmaker operates in the stationary equilibrium, and that investors can trade only when they meet the marketmaker. Since this marketmaker is monopolistic, he can make take-it-or-leave-it offers to investors. Then, the marketmaker would charge ΔV_{hk} to buyers of asset k (the ask price), and pay ΔV_{lk} to sellers of asset k (the bid price). In other words, the buyer's reservation value is the ask price, and the seller's reservation value is the bid price. Following this interpretation, condition (18) implies that an asset that is easier to find (with a larger μ_{lok}) has a narrower bid-ask spread. This suggests a negative relationship between liquidity for buyers and bid-ask spread.

Proposition 4 and 5 imply bid-ask spread decreases with μ_{lok}, but it does not have an unambiguous relationship with v_k. Thus, I conclude that bid-ask spread is a proxy for *illiquidity for buyers*: An asset with a larger bid-ask spread is harder to find.

Trading Volume Since all contacts between a buyer and a seller have a positive trade surplus, all contacts end with a trade. Then, we define trading volume of an asset as the total rate of contacts between buyers and sellers of that asset: $2zv_k\mu_{hn}\mu_{lok}$. Proposition 5 establishes the positive relationship between s_k and μ_{lok}. Moreover, Equation (32) implies $2zv_k\mu_{hn}$ is an increasing function of s_k. Combining these two facts, an asset with higher supply has a higher trading volume.

By (32), trading volume of asset k is equal to $\gamma_d s_k - (\gamma_d + \gamma_u)\mu_{lok}$. Since μ_{lok} is negatively related to σ_k, trading volume is positively related to σ_k. In other words, an asset with a higher payoff volatility has a higher trading volume. This is because the effect of risk on demand side ($2zv_k\mu_{hn}$) is dominant over the effect of risk on supply side (μ_{lok}). Volatility has a negative impact on liquidity for buyers (decreases μ_{lok}), but it has a positive impact on liquidity for sellers (increases $2zv_k\mu_{hn}$). The net effect of risk on trading volume is positive. Thus, Proposition 4 and 5 imply trading volume is positively related to $2zv_k\mu_{hn}$, but it does not have an unambiguous relationship with μ_{lok}. Consequently, trading volume is a proxy for *liquidity for sellers*.

Turnover In investment jargon, turnover is defined as the volume of shares traded during a particular period, as a fraction of total shares listed. Then, in our model turnover becomes $\frac{2zv_k\mu_{hn}\mu_{lok}}{s_k}$. By (32), turnover of asset k is equal to $\gamma_d - (\gamma_d + \gamma_u)\frac{\mu_{lok}}{s_k}$. Since $\frac{\partial \mu_{lok}}{\partial s_k} < 1$, turnover is an increasing function of s_k.

Since μ_{lok} is negatively related to σ_k, the above equation implies turnover is positively related to σ_k. The intuition is same as for trading volume: Volatility has a negative impact on liquidity for buyers (decreases μ_{lok}), but it has a positive impact on liquidity for sellers (increases $2zv_k\mu_{hn}$). The net effect of risk on turnover is positive. Thus, Proposition 4 and 5 imply turnover is positively related to $2zv_k\mu_{hn}$, but it does not have an unambiguous relationship with μ_{lok}. Consequently, turnover is a proxy for *liquidity for sellers*.

Market Capitalization In investment jargon, market capitalization is defined as the total value of the issued shares. Then, in my model market capitalization becomes $p_k s_k$. Since an asset with higher supply is sold at a higher price, market capitalization becomes an increasing function of s_k. On the other hand, market capitalization becomes a decreasing function of σ_k since an asset with higher payoff volatility is sold at a lower price. Consequently, market capitalization is a proxy for *liquidity for buyers* because, by Proposition 4 and 5, it comoves with μ_{lok} but its relation with $2zv_k\mu_{hn}$ is ambiguous.

4.2 Flight-to-liquidity

In this section, I will present a slight variation of the model to observe the effect of uncertainty in the market on liquidity premia between assets. Since liquidity premium is the focus of analysis, I should control for risk and obtain a cross-sectional variation in asset returns which is exclusively due to liquidity differences. To this end, I assume all assets have the same payoff volatility σ_s. Then, cross-sectional differences between asset prices will be pure liquidity premia. I use index s because I will vary the volatility to make a comparative statics analysis of flight-to-liquidity.

The related indifference condition for this model becomes

$$\frac{r + (1-q)(\gamma_d + \gamma_u)}{(1-q)x} \frac{1}{2z\mu_{lok}} + \frac{2z\gamma_d s_k q}{(1-q)x} \frac{1}{(2z\mu_{lok})^2} + \frac{1}{x} = \frac{\sigma_s}{W_s} \tag{34}$$

This quadratic equation allows one to write $2z\mu_{lok} = n_k(\frac{W_s}{\sigma_s})$, for some $\frac{W_s}{\sigma_s} < x$, and for some continuous and increasing function $n_k(.)$. After some algebra, I get

$$2z(\frac{\gamma_u}{\gamma_d+\gamma_u}-S)+\sum_{k=1}^{K}n_k(\frac{W_s}{\sigma_s})-2z\gamma_d\sum_{k=1}^{K}\frac{s_k}{n_k(\frac{W_s}{\sigma_s})}+K(\gamma_d+\gamma_u)=0 \tag{35}$$

The left-hand side of (35) is strictly increasing in $\frac{W_s}{\sigma_s}$ because $n_k(.)$ is strictly increasing for each k. Hence, (35) uniquely characterizes a candidate equilibrium $\frac{W_s}{\sigma_s}$. Existence is established by Proposition 3, since this model is a special case of my general model. Thus, I found that there is a unique $\frac{W_s}{\sigma_s}$ for any vector of asset supplies. Then, for a fixed vector $(s_1, ..., s_K)$ of asset supplies, one can write $W_s = T\sigma_s$ for some $T > 0$. This implies the flow value W of searching for an asset is higher in an equilibrium with high volatility compared to an equilibrium with low volatility.

Since $\frac{W_s}{\sigma_s}$ is unique, Equation (34) implies that the distribution of types is also unique i.e. independent of volatility level. Intuition is the following: In equilibrium, return differences between assets are liquidity premia caused by exogenous differences between asset supplies. Then, buyers choose their trading specialists allocation according to differences in asset supplies. Consequently, this gives us a unique trading specialists allocation. Proposition 1, in turn, implies a unique distribution of types. $W_s \equiv 2z\mu_{lok}(1-q)(\Delta V_{hk,s} - \Delta V_{lk,s})$ implies $\Delta V_{hk,s} - \Delta V_{lk,s}$ is an increasing function of σ_s for all $k \in \{1, ..., K\}$. In other words, bid-ask spreads are higher in an equilibrium with high volatility compared to an equilibrium with low volatility.

The pricing equation for this model is

$$p_{k,s} = \frac{\delta - A\sigma_s^2}{r} - \frac{W_s}{r} - (1+\frac{\gamma_d}{r(1-q)})\frac{W_s}{2z\mu_{lok}} \tag{36}$$

Since W_s increases with σ_s, Equation (36) implies prices are lower in an equilibrium with high volatility compared to an equilibrium with low volatility.

Let $s_1 > s_2$. Then, $\mu_{lo1} - \mu_{lo2} > 0$ and $p_{1,s} - p_{2,s} > 0$ by Proposition 4 i.e. an asset with more tradeable shares is sold at a higher price. This difference is equal to

$$p_{1,s} - p_{2,s} = (\frac{1}{2z\mu_{lo2}} + \frac{\gamma_d}{2z(1-q)\mu_{lo2}} - \frac{1}{2z\mu_{lo1}} - \frac{\gamma_d}{2z(1-q)\mu_{lo1}})W_s \tag{37}$$

The terms in the parenthesis in right-hand side are independent of the volatility level. Thus, $p_{1,s} - p_{2,s} = YW_s$ for some $Y > 0$. Let $R_{2,s} - R_{1,s}$ be the liquidity premium between asset 1 and asset 2. Then, it is equal to $\frac{\delta(p_{1,s} - p_{2,s})}{p_{1,s}p_{2,s}}$. Since the numerator increases with σ_s (because $p_{1,s} - p_{2,s} = YW_s$ and W_s increases with σ_s) and the denominator decreases with σ_s, liquidity premia are higher in an equilibrium with high volatility compared to an equilibrium with low volatility. The above discussion can be summarized in the following proposition:

Proposition 6 (Flight-to-liquidity) *In equilibrium, $\sigma_H > \sigma_L$ implies*

$\Delta V_{hk,H} - \Delta V_{lk,H} > \Delta V_{hk,L} - \Delta V_{lk,L}$, $\mu_{lok,H} = \mu_{lok;L}$, $p_{k,H} < p_{k,L}$ *for all* $k \in \{1, ..., K\}$ *and*

$R_{j,H} - R_{i,H} > R_{j,L} - R_{i,L}$ for all $i \neq j \in \{1, ..., K\}$ with $s_i > s_j$.

In words, when there is more uncertainty in the market, bid-ask spreads and liquidity premia are higher and prices are lower. This relationship between uncertainty and liquidity premia is empirically documented by Amihud and Mendelson (1991), Kamara (1994), Strebualev (2003) and Longstaff (2004b).

The next thing I do is to study the relationship between returns and liquidity factors for a "random" cross section of 200 assets. I compute two steady state equilibria of the theoretical model: In the first one, uncertainty is lower than the other equilibrium: $\sigma_1 < \sigma_2$. This numerical exercise suggests that the predictions of theoretical model developed in this paper are qualitatively consistent with much of the evidence from the empirical literature on flight-to-liquidity.

Both equilibria are computed for the same randomly generated economy of $K = 200$

asset types. The asset supplies s_k are drawn independently from an uniform distribution on some interval $[\underline{s}, \overline{s}]$. The expected flow payoff δ is set to 1. The bargaining powers $q_1, ..., q_K$ of sellers of assets $1, ..., K$ are drawn independently from an uniform distribution on some interval $[\underline{q}, \overline{q}]$. Varying q across trading pairs is a simple way to check the robustness of the results to the introduction of other forms of asset heterogeneity. The equilibrium expected return $R_{k,s} = \delta / p_{k,s}$ is plotted against various measures of liquidity used in the empirical literature, which have direct counterparts in the theoretical model. The relative bid-ask spread is $1 - \frac{\Delta V_{lk,s}}{\Delta V_{hk,s}}$. The trading volume is $2zv_k\mu_{hn}\mu_{lok}$. The turnover is $\frac{2zv_k\mu_{hn}\mu_{lok}}{s_k}$. The market capitalization is $p_k s_{k.}$ The values of the exogenous parameters are as in Table 1.

Parameters		**Value**
Contact Intensity	z	12000
Intensity of Switching to High	γ_u	1
Intensity of Switching to Low	γ_d	0.1
Discount Rate	r	4%
Number of Assets	K	200
Asset Supplies	s_k	$Uniform([510^{-4}/100, 1.510^{-3}/100])$
Expected Payoff	δ	1
Bargaining Power	q_k	$Uniform([0.45, 0.55])$
Hedging Cost of Low Type	x	2.4
Cost of Bearing Risk	A	18
Payoff Volatilities	σ_s	$\{13.74\%, 16.53\%\}$

Table 1: Parameter Values used in Comparison of Steady-states

The unit of time is one year. Assuming that the stock market opens 250 days a year and that there are 10 trading hours per day, $z = 12,000$ means that an investor establishes a contact every 12.5 minutes, on average. Given the chosen uniform distribution for s_k, the expected aggregate supply of assets, $E(\sum_{k=1}^{K} s_k)$, is 0.3. As in Duffie et al. (2005), an investor has a low marginal utility, on average, for 1 year out of every 11 years.

The discount rate r is set to 4%, consistent with Ibbotson's (2004) average T-bill rate of 3.8% during the period from 1926 to 2002. I select x and A based on assets' risk premia, measured by the difference $\frac{\delta}{p_k} - r$ between expected returns and the riskless rate. For $x = 0.4$

and $A = 18$, risk premia are about 2% in low volatility equilibrium and about 4% in high volatility equilibrium.

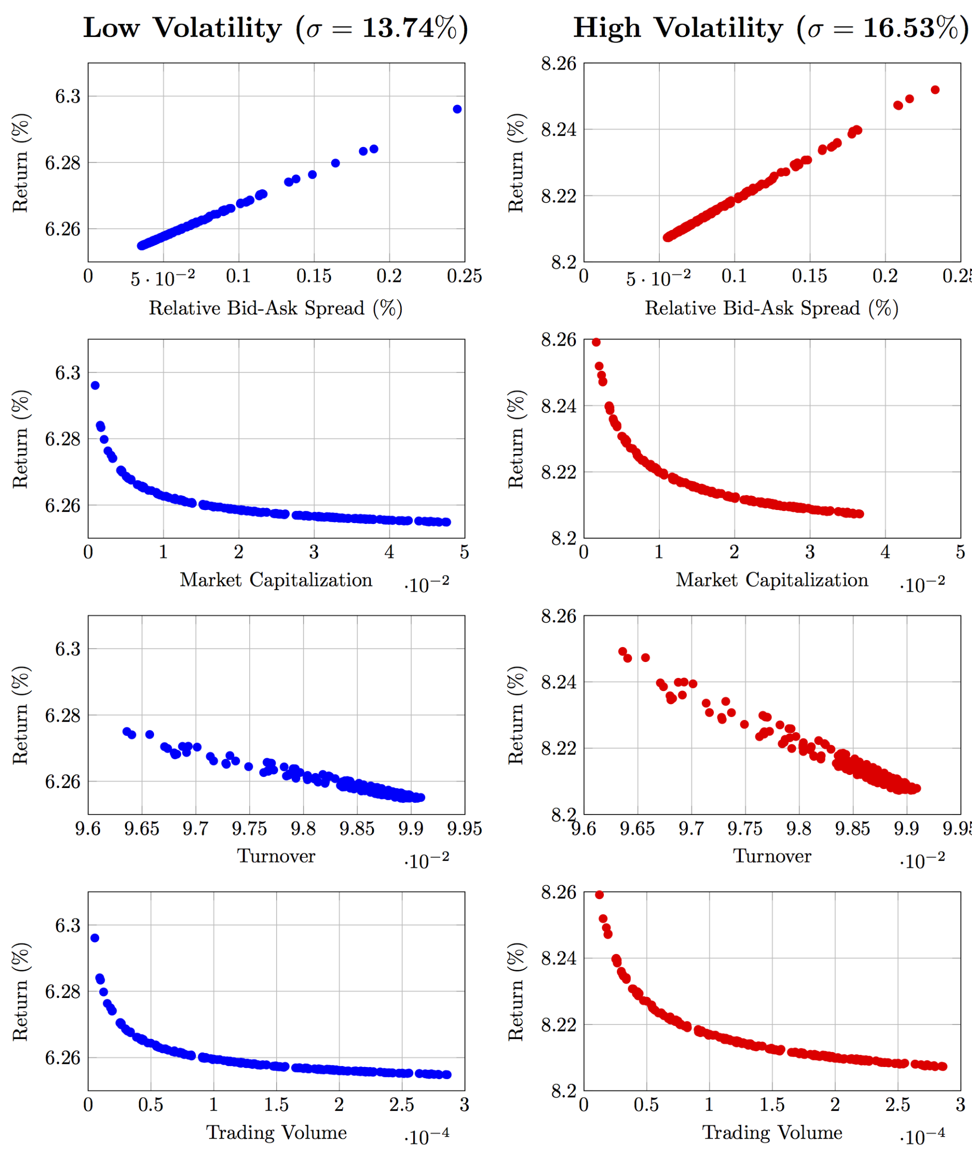

Figure 1. Cross-sectional variation in returns, explained by "liquidity factors"

Figure 1 displays the results of the computations. Returns and relative bid-ask spreads are positively related. In contrast with the theoretical results of Amihud and Mendelson (1986), the relationship is almost linear and not concave. Consistently with the empirical evidence, returns are negatively related to turnover and trading volume.

The sensitivity of returns to liquidity measures is higher in the high volatility equilibrium. That is, when there is more uncertainty in the market, liquidity differences create larger differences in returns suggesting that times of high volatility are associated with a flight to liquidity.

5 Conclusion

This paper contributes to a recent literature, spurred by Duffie et al. (2005), by presenting a model of a dynamic bargaining market that operates via search and matching in the spirit of Weill (2008). I complement this literature by treating risk-averse investors and multiple assets at the same time. Unlike the existing body of work in this literature, the uncertainty of asset payoffs is a factor of liquidity, which in turn opens the door to many interesting results, such as flight-to-liquidity. Theoretical and numerical results show that the model generates key qualitative facts documented in the empirical literature. Further work might extend the current model to incorporate stochastic variation in aggregate volatility.

Appendix A. Microfoundations for the risk-adjusted utility

Duffie et al. (2007) and Vayanos and Weill (2008) also provide a formal model of the holding cost. We model an economy with two heterogeneous assets, while Duffie et al. (2007) model an economy with only one asset and Vayanos and Weill (2008) model an economy with two identical assets.

Investors can invest in a riskless asset with return r and in two risky assets paying the cash flows given below. Cash flows are described by the cumulative dividend processes

$$dD_{kt} = \delta_k dt + \sigma_k dB_t$$

for $k \in \{1, 2\}$, where δ and σ_k are positive constants, and B_t is a standard Brownian motion. Investors derive utility from the consumption of a numéraire good, and have a CARA utility function

$$-E[\int_0^\infty \exp(-\alpha c_t - \beta t)dt] \tag{38}$$

Each investor receives a cumulative endowment process

$$de_t = \sigma_e[\rho_t dB_t + \sqrt{1 - \rho_t^2} dZ_t]$$

where σ_e is a positive constant, Z_t a standard Brownian motion independent of B_t, and ρ_t the instantaneous correlation between the dividend process and the endowment process. The process ρ_t can take two values: $\rho_t = 0$ for high-type investors and $\rho_t = \underline{\rho} > 0$ for low-type investors. The processes (ρ_t, Z_t) are pairwise independent across investors. We set $A \equiv r\alpha/2$ and $x \equiv r\alpha\underline{\rho}\sigma_e$.

An investor maximizes (38) subject to the budget constraint

$$dW_t = [rW_t - c_t + \sum_{i=1}^{2}(\delta_i - rp_i)q_{it}]dt + [\sum_{i=1}^{2}\sigma_i q_{it} + \rho_t\sigma_e]dB_t + \sigma_e\sqrt{1-\rho_t^2}dZ_t \tag{39}$$

and the transversality condition

$$\lim_{T\to\infty} E[\exp(-r\alpha W_T - \beta T)] = 0 \tag{40}$$

where W_t is the wealth and q_{it} is the number of shares invested in asset $i \in \{1,2\}$. The investor's controls are the consumption $c \in \mathbb{R}$ and the investments $(q_1, q_2) \in \mathbb{Z}^2$. To derive the optimal rules, the technique of stochastic dynamic programming is used following Merton (1971). Define

$$J(\rho_t, W_t, t) \equiv \max_{\{c_t, q_{1t}, q_{2t}\}} - E_t[\int_t^\infty \exp(-\alpha c_s - \beta s)ds]$$

subject to (39) and (40). Then

$$\begin{aligned}
J(\rho_t, W_t, t) &= \max_{\{c_t, q_{1t}, q_{2t}\}} - E_t[\int_t^{t+\Delta t} \exp(-\alpha c_s - \beta s)ds + \int_{t+\Delta t}^\infty \exp(-\alpha c_s - \beta s)ds] \\
&= \max_{\{c_t, q_{1t}, q_{2t}\}} - \exp(-\alpha c_t - \beta t)dt + E_t\{\max_{\{c_t, q_{1t}, q_{2t}\}} - E_{t+\Delta t}[\int_{t+\Delta t}^\infty \exp(-\alpha c_s - \beta s)ds]\} \\
&= \max_{\{c_t, q_{1t}, q_{2t}\}} - \exp(-\alpha c_t - \beta t)dt + E_t[J(\rho_{t+\Delta t}, W_{t+\Delta t}, t+\Delta t)] \\
&= \max_{\{c_t, q_{1t}, q_{2t}\}} - \exp(-\alpha c_t - \beta t)dt + E_t[J(\rho_t, W_t, t) + \frac{dJ}{dt}]
\end{aligned}$$

in the limit as $\Delta \to 0$, subject to (39) and (40). By Ito's lemma,

$$E[dJ] = J_W E[dW] + J_t dt + \frac{1}{2}J_{WW}E[(dW)^2] + \gamma(\rho)[J(\rho', W, t) - J(\rho, W, t)]$$

where $\gamma(\rho) = \gamma_u$ if $\rho = \underline{\rho}$ and $\gamma(\rho) = \gamma_d$ if $\rho = 0$. Thus, the investor's problem becomes

$$0 = \max_{c,q_1,q_2} - \exp(-\alpha c_t - \beta t)dt + J_W E[dW] + J_t dt + \frac{1}{2} J_{WW} E[(dW)^2] + \gamma(\rho)[J(\rho', W, t) - J(\rho, W, t)]$$

where

$$E[dW] = [rW - c + \sum_{i=1}^{2} (\delta_i - rp_i) q_i] dt$$

$$E[(dW)^2] = ((\sum_{i=1}^{2} \sigma_i q_i)^2 + 2\rho\sigma_e \sum_{i=1}^{2} \sigma_i q_i + \sigma_e^2) dt$$

and $\gamma(\rho) = \gamma_u$ if $\rho = \underline{\rho}$ and $\gamma(\rho) = \gamma_d$ if $\rho = 0$. Hence, suppressing the time argument t, the investor's value function $J(\rho, W)$ satisfies the HJB equation

$$0 = \sup_{c,q_1,q_2} \{- \exp(-\alpha c) + D^{(c,q)} J(\rho, W) - \beta J(\rho, W)\} \tag{41}$$

where

$$\begin{aligned} D^{(c,q)} J(\rho, W) &= J_W(\rho, W)[rW - c + \sum_{i=1}^{2} (\delta_i - rp_i) q_i] \\ &+ \frac{1}{2} J_{WW}(\rho, W)((\sum_{i=1}^{2} \sigma_i q_i)^2 + 2\rho\sigma_e \sum_{i=1}^{2} \sigma_i q_i + \sigma_e^2) + \gamma(\rho)[J(\rho', W) - J(\rho, W)] \end{aligned}$$

and where the transition intensity $\gamma(\rho) = \gamma_u$ for $\rho = \underline{\rho}$ and $\gamma(\rho) = \gamma_d$ for $\rho = 0$. We guess that $J(\rho, W)$ takes the form

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

for some function $V(\rho)$. Replacing into (41), we find that the optimal consumption is

$$c(\rho, W) = -r[W + V(\rho)] + \frac{r - \beta + \frac{r^2\alpha^2\sigma_e^2}{2}}{r\alpha}$$

and the optimal investment satisfies

$$q(\rho) \in \arg\max_{q_1,q_2} \{C(\rho, q_1, q_2) - rp_1q_1 - rp_2q_2\} \tag{42}$$

where $C(\rho, q_1, q_2)$ is the incremental certainty equivalent of holding q_1 shares of asset 1 and q_2 shares of asset 2 relative to holding none. Using the definitions of A and x, we can write the certainty equivalents as $C(0, q_1, q_2) = \sum_{i=1}^{2} \delta_i q_i - A(\sum_{i=1}^{2} \sigma_i q_i)^2$ for high-type investors and $C(\underline{\rho}, q_1, q_2) = \sum_{i=1}^{2} \delta_i q_i - A(\sum_{i=1}^{2} \sigma_i q_i)^2 - x\sum_{i=1}^{2} \sigma_i q_i$ for low-type investors.

Plugging $c(\rho, W)$ back into (41), we find that (41) is satisfied iff

$$0 = -rV(\rho) + \max_{q_1,q_2}\{C(\rho, q_1, q_2) - rp_1q_1 - rp_2q_2\} + \gamma(\rho)\frac{1 - \exp(-A(V(\rho') - V(\rho))}{A} \tag{43}$$

By (43), we get a system of two equations in two unknowns $V(0)$ and $V(\underline{\rho})$, and it is easy to check that it has a unique solution. Since investors are allowed to hold only zero or one unit of some asset in our main model, the relevant certainty equivalents are:

$C(0, 1, 0) = \delta_1 - A\sigma_1^2$

$C(\underline{\rho}, 1, 0) = \delta_1 - x\sigma_1 - A\sigma_1^2$

$C(0, 0, 1) = \delta_2 - A\sigma_2^2$

$C(\underline{\rho}, 0, 1) = \delta_2 - x\sigma_2 - A\sigma_2^2$.

Due to our assumption on asset holdings, these certainty equivalents apply to an economy with any number of assets. As long as investors are allowed to invest in only one of the risky assets, cross-asset terms in (42) cancel, and certainty equivalent flows described above apply without loss of generality.

References

[1] Acharya, V. V., & Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, 77, 375–410.

[2] Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17, 223-249.

[3] Amihud, Y., & Mendelson, H. (1991). Liquidity, maturity, and the yield on US Treasury securities. *Journal of Finance*, 46, 479-486.

[4] Barinov, A. (2014). Turnover: liquidity or uncertainty? *Management Science, 60*(10), 2478-2495.

[5] Chordia, T., Roll, R., & Subrahmanyam, A. (2000). Commonality in liquidity. *Journal of Financial Economics*, 56, 3–28.

[6] Comiskey, E. E., Walkling, R. A., & Weeks, M. (1987). Dispersion of expectations and trading volume. *Journal of Business Finance & Accounting,* 14, 229-239.

[7] Constantinides, G. M. (1986). Capital market equilibrium with transaction costs. *Journal of Political Economy*, 94, 842–862.

[8] Diamond, P. (1982). Aggregate demand management in search equilibrium. *Journal of Political Economy*, 90, 881–94.

[9] Duffie, D., Garleanu, N., & Pedersen, L. H. (2002). Securities lending, shorting, and pricing. *Journal of Financial Economics*, 66, 307–39.

[10] Duffie, D., Garleanu, N., & Pedersen, L. H. (2005). Over-the-counter markets. *Econometrica*, 73, 1815–1847.

[11] Duffie, D., Garleanu, N., & Pedersen, L. H. (2007). Valuation in over-the-counter markets. *Review of Financial Studies*, 20, 1865–1900.

[12] Duffie, D., & Sun, Y. (2007). Existence of independent random matching. *Annals of Applied Probability*, 17, 387–419.

[13] Gallant, A. R., Rossi, P. E., & Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies, 5*(2), 199-242.

[14] Garleanu, N. (2006). Portfolio choice and pricing in imperfect markets. *Mimeo.*

[15] Giroux, G. (2005). Markets of a large number of interacting agents. *Mimeo.*

[16] Huang, M. (2003). Liquidity shocks and equilibrium liquidity premia. *Journal of Economic Theory*, 109, 104–29.

[17] Ibbotson. (2004). *Stock, bonds, bills, and inflation statistical yearbook.* Chicago, IL: Ibbotson Associates.

[18] Judd, K. L. (1998). *Numerical methods in economics.* Boston, MA: MIT Press.

[19] Kamara, A. (1994). Liquidity, taxes, and short-term Treasury yields. *Journal of Financial and Quantitative Analysis*, 29, 403-417.

[20] Karpoff, J. (1987). The relation between price changes and trading volume: a survey. *Journal of Financial and Quantitative Analysis*, 22, 109-126.

[21] Krainer, J., & LeRoy, S. (2002). Equilibrium valuation of illiquid assets. *Economic Theory*, 19, 223–42.

[22] Lang, L. H. P., Litzenberger, R. H., & Madrigal, V. (1992). Testing financial market Equilibrium under asymmetric information. *Journal of Political Economy, 100*(2), 317-48.

[23] Li, J., & Wu, C. (2006). Daily return volatility, bid-ask spreads, and information flow: analyzing the information content of volume. *Journal of Business*, 79, 2697-2739.

[24] Longstaff, F. (2004a). Financial claustrophobia: asset pricing in illiquid markets. *Mimeo.*

[25] Longstaff, F. (2004b). The flight-to-liquidity premium in US Treasury bond prices. *Journal of Business*, 77, 511-526.

[26] Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory, 3*(4), 373-413.

[27] Miao, J. (2006). A search model of centralized and decentralized trade. *Review of Economic Dynamics*, 9, 68–92.

[28] Pastor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111, 642-685.

[29] Protter, P. (2004). *Stochastic integration and differential equations.* New York, NY: Springer.

[30] Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance, 44*(5), 1115-1153.

[31] Strebulaev, I. (2003). Liquidity and asset pricing: evidence from the US Treasury securities market. *Mimeo.*

[32] Taylor, A. E., & Mann, R. W. (1983). *Advanced calculus.* New York, NY: Wiley, John and Sons.

[33] Trejos, A., & Wright, R. (1995). Search, bargaining, money, and prices. *Journal of Political Economy*, 103, 118–40.

[34] Vayanos, D. (1998). Transaction costs and asset prices: a dynamic equilibrium model. *Review of Financial Studies,* 11, 1–58.

[35] Vayanos, D. (2004). Flight to quality, flight to liquidity, and the pricing of risk. *Mimeo.*

[36] Vayanos, D., & Wang, T. (2007). Search and endogenous concentration of liquidity in asset markets. *Journal of Economic Theory*, 136, 66 – 104.

[37] Vayanos, D., & Weill, P.-O. (2008). A search-based theory of the on-the-run phenomenon. *Journal of Finance*, 63, 1361-1398.

[38] Weill, P.-O. (2007). Leaning against the wind. *Review of Economic Studies*, 74, 1329-1354.

[39] Weill, P.-O. (2008). Liquidity premia in dynamic bargaining markets. *Journal of Economic Theory*, 140, 66-96.