working paper

06 07

# Search in Asset Markets

by Ricardo Lagos and Guillaume Rocheteau

FEDERAL RESERVE BANK OF CLEVELAND

**Working Paper 06-07** July 2006

# Search in Asset Markets

by Ricardo Lagos and Guillaume Rocheteau

We investigate how trading frictions in asset markets affect portfolio choices, asset prices and efficiency. We generalize the search-theoretic model of financial intermediation of Duffie, Gârleanu and Pedersen (2005) to allow for more general preferences and idiosyncratic shock structure, unrestricted portfolio choices, aggregate uncertainty and entry of dealers. With a fixed measure of dealers, we show that a steady-state equilibrium exists and is unique, and provide a condition on preferences under which a reduction in trading frictions leads to an increase in the price of the asset. We also analyze the effects of trading frictions on bid-ask spreads, trade volume and the volatility of asset prices, and find that the asset allocation is constrained-inefficient unless investors have all the bargaining power in bilateral negotiations with dealers. We show that the dealers' entry decision introduces a feedback that can give rise to multiple equilibria, and that free-entry equilibria are generically inefficient.

*Keywords*: asset prices, bid-ask spread, execution delay, liquidity, search, trade volume
*JEL Classification*: G11, G12, G21

Ricardo Lagos is at the Federal Reserve Bank of Minneapolis and New York University. He can be reached at ricardo.lagos@nyu.edu. Guillaume Rocheteau is at the Federal Reserve Bank of Cleveland. He can be reached at guillaume.g.rocheteau@clev.frb.org. The first version of this paper was written in October 2005. The authors are grateful to Pierre-Olivier Weill and Nicolae Gârleanu for many helpful conversations. They also thank seminar participants at the Federal Reserve Bank of Cleveland, Indiana University, Singapore Management University, and the University of Pennsylvania for their comments. Financial support from the C.V. Starr Center for Applied Economics at NYU is gratefully acknowledged. Much of this research was conducted while Lagos was at the Minneapolis Fed and Rocheteau was visiting Penn.

# 1 Introduction

Trading takes time. The trading delays (or "frictions") which are at the foreground of the economic modeling of many markets, such as labor markets, are also present in asset markets. Even in equity and bond markets, where the trading arrangements are well developed, executing a trade requires resources—*time* being one of them. An investor wishing to rebalance his portfolio will usually contact a broker. In order to get the order executed, the broker will route it to an organized exchange, to a dealer or market-maker, or to some electronic trading system. Regardless of the method chosen, an order cannot be filled until a counterpart for the trade is found, and finding a counterpart takes time.[1]

Markets for financial securities have traditionally been the realm of the competitive Walrasian paradigm. Accordingly, trade in these markets, e.g., the matching of buyers to sellers, is typically regarded as an instantaneous and costless process—and left unmodeled. In this paper we further the view that trading frictions and the mechanics of trade are important for understanding the functioning of asset markets, including many of those for financial securities—and especially fixed-income securities—which are typically traded in a decentralized manner in over-the-counter markets.

Financial markets are currently in the midst of a technological revolution that is poised to reshape their structure and the ways in which they operate. The new electronic trading systems being introduced allow brokers and dealers to match their orders faster. For example, the advent of electronic communication networks (ECNs)—private electronic screen-based trading systems built around computer algorithms that match buy and sell orders through an open limit-order book—is allowing investors to find trading opportunities more rapidly, and sometimes even directly, without the need for a traditional intermediary. Since traders now have the option to submit orders to an ECN and trade with others directly instead of having their

[1] The time it takes to execute an order in an organized exchange such as the New York Stock Exchange (NYSE) or the National Association of Securities Dealers Automated Quotation System (NASDAQ) can range from a few seconds to several minutes. These delays vary considerably across market centers; see Boehmer (2005, Table 6) and Boehmer, Jennings and Wei (2005, Table 2). To an outsider, differences in execution delays of a few minutes may seem immaterial. But in a high-frequency marketplace where profit opportunities come and go very fast, a few minutes, or even seconds, can make a big difference to some traders. Boehmer (2005, Table 4, Panel A) reports that the median effective spread is 2.2 cents (0.09 percentage points of the price) larger in NASDAQ than in the NYSE, but that on average, trades execute 12.2 seconds faster in NASDAQ. This seems to suggest that traders are willing to pay a significant amount to get a trade executed just a few seconds faster. Trading delays tend to be longer—ranging from a few minutes to a day—for fixed-income securities traded in over-the-counter markets; see Schultz (2001) and Saunders, Srinivasan and Walter (2002).

trades intermediated by dealers, widespread use of these new technologies can drastically reduce intermediation and transaction costs and accelerate trade execution.[2] These large transformations in trading arrangements are bound to manifest themselves in asset prices and portfolio allocations and change the nature of financial intermediation, but how exactly?

In this paper we develop an equilibrium search-theoretic model of the exchange process in an asset market and use it to study how the degree of trading frictions—as determined by the number of financial intermediaries and available technologies—affects the overall performance of the market. We focus on the effects that these frictions have on portfolio allocations, intermediation costs (e.g., bid-ask spreads, intermediation fees), the level and volatility of asset prices, various measures of liquidity (e.g., execution delays, trade volumes) and allocative efficiency.

We build on the recent work of Duffie, Gârleanu and Pedersen (2005), and similarly to their work, our description of the asset market captures some salient features of over-the-counter markets and telephone-dealer markets such as those for commercial paper and corporate bonds. From a methodological point of view, we generalize the model in Duffie, Gârleanu and Pedersen (2005) to allow for more general preferences and structure of idiosyncratic shocks, aggregate uncertainty, unrestricted portfolio choices and entry of intermediaries. Keeping with the spirit of the advent of electronic trading systems, in addition we allow investors to sometimes trade directly with "the market," without requiring the services of an intermediary.

The paper is structured as follows. In Section 2 we lay out the basic model. Investors receive idiosyncratic preference shocks that change their desired portfolios and are able to rebalance their portfolios at random times. These random delays are meant to capture the order-execution delays that investors or dealers experience while they try to find a counterpart for the trade. In contrast to investors, dealers have instantaneous access to a competitive interdealer market. In intermediated trades, intermediation fees (or bid and ask prices) and portfolio allocations are determined jointly through bargaining in a bilateral match between an investor and a dealer.

[2]These new technologies are already having a big impact in equity markets. For example, more than 50 percent of all NASDAQ trades (about 26 percent of the dollar trade volume in the year 2000 according to NASDAQ, 2000) are now executed through ECNs. These technological innovations have been accompanied by changes in regulations that promote competition among intermediaries. For instance, the New Order Handling Rules introduced by the Securities Exchange Commission (SEC) in 1997 impose that the public orders posted on ECNs can compete more directly with NASDAQ market-makers. (See McAndrews and Stefanadis, 2000, for an account of the emergence of ECNs in U.S. equity markets and Barclay et al., 1999, pp. 4–7, for a detailed description of the SEC order-handling reforms.) The number of electronic trading systems for fixed-income securities has also proliferated from 11 in 1997 to 70 in 2000. It is estimated that in the year 2000, 40 percent of U.S. Treasury securities transactions were done electronically, up twofold from 1999 (see Allen, Hawkins and Sato, 2001).

Alternatively, when an investor gains direct access to the asset market, he chooses his new portfolio taking the asset price parametrically and pays no intermediation fee.

In Section 3 we study the model with a constant measure of dealers and show that a steady-state equilibrium exists and is unique. In Section 3.1 we extend the model to allow for aggregate uncertainty. Section 3.2 presents a calibrated version of the model used to illustrate and complement our analytical results.

Section 4 deals with the effects of trading frictions on asset prices. In Section 4.1 we show that for CRRA preferences, the price of the asset increases (decreases) as trading frictions are reduced if the coefficient of relative risk aversion is smaller (larger) than one. For a logarithmic utility function, the asset price is independent of trading frictions and corresponds to the price that would prevail in a frictionless Walrasian market. In Section 4.2 we use the formulation with aggregate uncertainty to study the effects of trading frictions on asset-price volatility.

Section 5 deals with the effects that the degree of trading frictions has on the volume of trade and the allocation of portfolios across investors. In Section 6 we show that intermediation fees are trade-specific and increase with the size of the portfolio reallocations. For a given trade size, fees increase with dealers' bargaining power and decrease with the frequency at which investors can access the market directly, as well as with the frequency at which they can meet dealers. Interestingly, since a reduction in the trading frictions increases the size of each trade, faster trading does not necessarily reduce the dealers' average profit.

In Section 7 we endogenize trading delays by allowing entry of dealers. We show that an equilibrium exists, but find that it need not be unique. We provide examples where multiple (e.g., three) steady states exist. This multiplicity arises because of a strategic complementarity between the investors' portfolio choices and the dealers' entry decisions. When multiple equilibria exist, some exhibit narrow bid-ask spreads, large traded volumes and short execution delays, while others display wide spreads, small volumes and long delays. We show that a reduction in trading frictions can remove the multiplicity. Thus, perhaps counter to intuition, it is possible that a regulatory reform or a technological innovation that gives investors more direct access to the asset market (such as ECNs) leads to a relatively large increase in market liquidity and results in a higher volume of intermediated trades.[3]

[3] Concerns have been raised that increased competition from alternative trading networks could reduce dealers' incentives to make markets, and hence adversely affect the liquidity of the market. Since the growth of electronic trading platforms is a very recent phenomenon, there is only a handful of academic studies that examine their effects. Weston (2000, 2002), for instance, finds that the increase in competition resulting from the growth of

Finally, in Section 8 we carry out a normative analysis. We establish that portfolio decisions tend to be inefficient because of a holdup problem introduced by bargaining. Investors with high valuations tend to invest too little, while those with low valuations tend to invest too much. This inefficiency is eliminated when dealers have no market power. When there is entry of dealers, the measure of dealers is efficient if the bargaining power of dealers coincides with their contribution to the matching process. Therefore, in this context, the inefficiencies on the intensive margin (portfolio choices) and extensive margin (number of dealers) cannot be corrected simultaneously.

Our paper belongs to the new search-theoretic literature on financial markets that includes Duffie, Gârleanu and Pedersen (2005), Miao (2006), Rust and Hall (2003), Vayanos and Wang (2002), Weill (2005a,b) and Gârleanu (2005).[4] Relative to these papers, our methodological contributions consist of: (*i*) relaxing the portfolio restrictions,[5] (*ii*) allowing for more general preferences and more general forms of investor heterogeneity, (*iii*) allowing for aggregate uncertainty, (*iv*) allowing investors direct, as well as indirect (i.e., dealer-intermediated), access to a competitive interdealer market, and (*v*) endogenizing the degree of the trading frictions by endogenizing the measure of dealers.[6] These generalizations yield new insights, e.g., on the link between trading frictions and the level and volatility of asset prices, on the potential for multiple equilibria, on the efficiency implications of the degree of market power of dealers, and on the size of intermediation fees as a function of the trade-size and the degree of trading frictions. From an applied standpoint, the generalizations we develop also allow us to address new issues, such as the impact that the emergence and growth of alternative trading systems will have on spreads, trade volumes, the number of financial intermediaries and execution delays.

---

trading through ECNs in NASDAQ has resulted in larger trade volumes, tighter bid-ask spreads and net entry of market-makers. See Weston (2002) for references to related work.

[4] Also related is the work of Spulber (1996), who considers a search environment where middlemen intermediate trade between heterogeneous buyers and sellers. There is also a large related literature, not search-based, which studies how exogenously specified transaction costs affect the functioning of asset markets. Recent examples include He and Modest (1995), Lo, Mamaysky and Wang (2004), Luttmer (1996) and Vayanos (1998), to name a few. See Heaton and Lucas (1995) for a survey of this body of work.

[5] As in the early search-theoretic models of money, e.g., Kiyotaki and Wright (1993), Duffie, Gârleanu and Pedersen (2005) restrict portfolio choices to lie in the set $\{0,1\}$. We impose no restrictions on the quantity of assets that an investor can hold except that it be nonnegative (i.e., we do not allow for short-selling of assets). Also as in Kiyotaki and Wright (1993), Duffie, Gârleanu and Pedersen (2005) allow for bilateral trades between investors. In our model, investors can trade directly with other investors every time they gain direct access to the competitive market—but these trades are not bilateral.

[6] Duffie, Gârleanu and Pedersen (2005) consider a search intensity decision by a monopolist dealership, while we model a market with many small dealers who compete with each other for order flow.

## 2 The environment

Time is continuous and the horizon infinite. There are two types of infinitely lived agents: a unit measure of investors and a unit measure of dealers. (We endogenize the measure of dealers in Section 7.) There is one asset and one perishable good, which we use as numeraire. The asset is durable and perfectly divisible. The stock of assets in the economy is $A \in \mathbb{R}_+$. The numeraire good is produced and consumed by all agents. The instantaneous utility function of an investor is $\mathcal{U}(a, c; i) = u_i(a) + c$, where $a \in \mathbb{R}_+$ represents the investor's asset holdings, $c \in \mathbb{R}$ is the net consumption of the numeraire good ($c < 0$ if the investor produces more of these goods than he consumes), and $i \in \{1, ..., I\}$ indexes a preference shock. The utility function $u_i(a)$ is continuously differentiable, strictly increasing and strictly concave.[7] Each investor receives a preference shock with Poisson arrival rate $\delta$. This process is independent across investors. Conditional on the preference shock, the probability the investor draws preference type $i$ is $\pi_i$, with $\sum_{i=1}^{I} \pi_i = 1$. These preference shocks capture the notion that investors will value the services provided by the asset differently over time, thereby generating a need for investors to reallocate their portfolios. Dealers cannot hold positions and their instantaneous utility is $c$, their consumption of the numeraire good.[8] All agents discount at rate $r > 0$.

There is a competitive market for the asset and dealers have continuous access to it. An investor can access this market indirectly through a dealer. Investors contact dealers at random according to a Poisson process with arrival rate $\alpha$. Once they have contacted each other, the dealer and the investor negotiate over the quantity of assets that the dealer will acquire for the investor and over the intermediation fee that the investor will pay the dealer for his services. After the transaction has been completed, the dealer and the investor part ways. Investors can also gain direct access to the competitive asset market according to an independent Poisson process with arrival rate $\beta$. The trading process is illustrated in Figure 1.

We regard this theoretical trading process as a stylized characterization of actual trading arrangements in some asset markets. We think of the investors in the model as representing investor-broker pairs who are searching for a counterpart to execute a given trade. A counter-

[7] One can think of the asset as being a durable good that provides a flow of services to its owner, such as a house or a car. Also, $a$ could be thought of as shares of a "tree" that yield a real, perishable "fruit" dividend different from the numeraire good. Alternatively, one may adopt the interpretation of Duffie, Gârleanu and Pedersen (2005) and consider $u_i(a)$ as a reduced-form utility function that stands for the various motives an investor may have for holding the asset, such as liquidity or hedging needs.

[8] The restriction that dealers cannot hold assets is of no consequence when analyzing steady-state equilibria, as we do in most of the paper. See Weill (2005a) for dynamic equilibria where dealers hold positions.

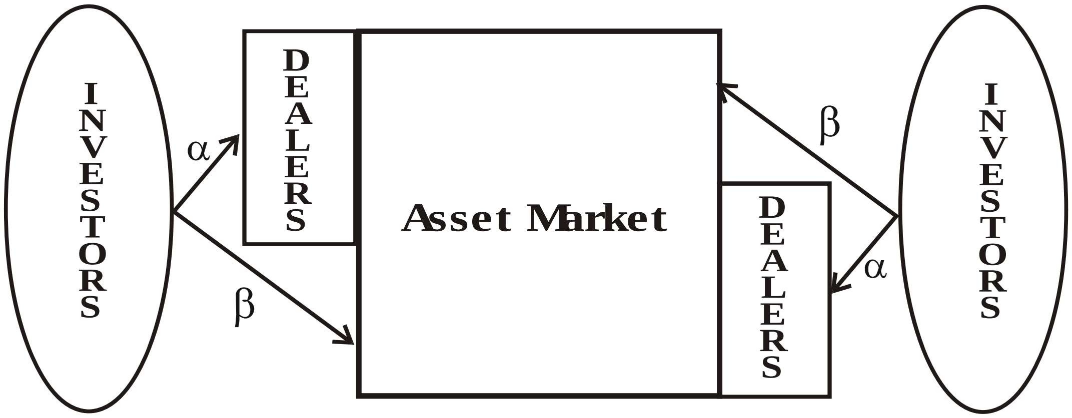


Figure 1: Trading process

part can typically be found in a traditional intermediary, such as a dealer or market-maker, or through electronic trading systems that allow investors to trade directly without an intermediary. Finding a counterpart by either method involves delays. In the model, these execution delays correspond to $1/\alpha$ and $1/\beta$. Bilateral trade and bargaining between investors/brokers and dealers is a feature of many financial transactions, both in quote-driven and order-driven markets. The fact that agents trade assets at a competitive price when they gain access to the market captures the idea that even though there are execution delays, investors and dealers sometimes interact in large groups and in these instances, they take prices as given.

### 2.1 Discussion

While our theoretical model is stylized, we believe it captures the salient features of many financial trades in various contexts. First, as Duffie, Gârleanu and Pedersen (2005) emphasize, the model incorporates the key elements of over-the-counter (OTC) markets. The defining feature of OTC markets is that they have no formal organization: they do not have a physical location and operate in a completely decentralized manner. A typical OTC market consists of brokers and dealers who can be located all over the country and negotiate directly with one another over computer networks and by telephone. Unlisted stocks, some derivatives and most debt instruments, such as commercial paper, corporate and municipal bonds, are examples of securities traded over the counter.

Trade in an OTC market is a textbook search problem: buyers and sellers seek each other out and usually trade in pairs. An investor wishing to trade an OTC security will contact, electronically or by phone, dealers who are specialized in that particular asset category. It typically takes a dealer several minutes to generate a quote. After the dealer produces the initial quote, negotiations will ensue regarding the price quoted and other details of the trade, in particular the quantity traded.[9] In the model, $1/\alpha$ has a natural interpretation as the average time it takes the investor/broker to find a dealer with whom they reach an agreement regarding the terms of the trade. Another key feature of OTC markets is that dealers have access to interdealer brokered networks where they can manage their inventories. In this spirit, dealers in our model have access to a competitive asset market where they can continuously manage their asset positions.[10]

In some cases, even listed securities may get traded OTC-style, i.e., in a decentralized dealer network, bilaterally, and with negotiated prices and quantities. Such is the case for large blocks of stocks and some treasuries. Although these securities may trade in organized exchanges for small sizes, brokers routinely have to find suitable counterparts for large blocks. The first problem that block traders face is searching for traders with these latent demands, and once they have found them, the terms of the transaction will be negotiated bilaterally.[11]

Bilateral trade with negotiated prices and quantities are also features of some transactions conducted in the context of organized exchanges. For instance, trade in quote-driven dealer-organized markets such as NASDAQ is also bilateral, either between two registered dealers—market-makers—or between an investor/broker and a market-maker. Regarding the terms of trade, market-makers are typically required to post prices at which they are willing to buy (the bid price) and sell (the ask price) some quantity of the instrument they are registered for. In principle, these price quotes can be *soft* or *firm*. Dealers who offer soft quotes can revise their prices when asked to trade, or even refuse to trade. Soft quotes lead to bargaining. Firm quotes are good only up to the quantity that the dealer specifies in the quote, so they may also coexist

[9] See Harris (2003) for examples and detailed accounts of real-world trades and Schultz (2001) or Saunders, Srinivasan and Walter (2002) for a description of over-the-counter corporate bond markets.

[10] Our competitive interdealer market allows dealers to intermediate transactions instantaneously without having to take positions, and hence, without having to assume inventory risk. In reality, in addition to intermediating between investors, some dealers—*position traders*—take positions in the hope of making capital gains. In this paper we abstract from this inventory risk. But there are also *pure spread traders*, i.e., dealers who—like the ones in our model—don't take positions, and profit exclusively from buying low and selling high.

[11] As an example, the London International Financial Futures and Options Exchange (LIFFE) introduced a block trading facility in April 1999. LIFFE's block trading procedures permit its members and their qualified clients to quickly trade large blocks at bilaterally negotiated prices.

with elements of bargaining.[12]

Nowadays, listed stocks also trade in various *alternative trading systems*—ECNs being the best known among these electronic exchanges. Accordingly, another key feature of modern financial exchanges that we try to capture in our formal modeling is that, for many securities, markets *overlap*. That is, when a broker receives a trade order for a particular security, he can route it to a traditional dealer or exchange, or to an alternative trading system. The direct access to the asset market that we grant investors in our formal modeling is meant to capture the trade opportunities created by these alternative trading systems that match investors directly. For example, $1/\beta$ in the model can be thought of as the average time it takes for a market order to fill when routed through an electronic trading system such as an ECN.[13]

# 3 Equilibrium

In this section we characterize stationary equilibria where the joint distribution of portfolios and preferences across investors remains constant over time. Consider an investor with a preference type $i$ who holds stock of assets $a$. The value function of the investor, $V_i(a)$, satisfies the following Bellman equation

$$\begin{aligned} rV_i(a) = u_i(a) + \delta \sum_k \pi_k \left[V_k(a) - V_i(a)\right] + \alpha[V_i(a_i^b) - V_i(a) - p(a_i^b - a) - \phi_i\,(a)] \\ + \beta \max_{a_i^c}[V_i(a_i^c) - V_i(a) - p(a_i^c - a)], \end{aligned} \tag{1}$$

for $a \in \mathbb{R}_+$ and $i = 1, ..., I$. According to (1), the investor enjoys a utility flow $u_i(a)$ from holding portfolio $a$. He receives a new preference shock with instantaneous probability $\delta$, and conditional on this shock, he draws a new preference type $k$ with probability $\pi_k$ and enjoys a capital gain $V_k(a) - V_i(a)$. With instantaneous probability $\alpha$, the investor meets a dealer who can help him rebalance his portfolio. Upon contacting a dealer, the investor buys $a_i^b - a$ (sells

[12] NASDAQ requires its dealers to quote firm two-sided markets, i.e., both a bid and an ask price, for at least 100 shares of the security they are registered for. Notwithstanding, it is not uncommon for dealers to trade at different prices than those they quote to the public. This happens because market-makers' posted prices are only binding up to the quantities specified in the quote, so a trader wishing to trade a larger quantity will end up negotiating the terms of the trade bilaterally with the dealer.

[13] In reality, the cost of sending an order to an ECN is usually small compared to the fees or spreads charged by traditional dealers. For this reason, the cost of trading directly in the asset market is normalized to zero in the model. Many interdealer markets are now open to investors (e.g., the trading platforms BrokerTec and E-Speed). Accordingly, in our formal model, the market where investors can trade directly without being intermediated by dealers is the same interdealer market where dealers rebalance their asset holdings. Nothing would change if we modeled them as being distinct marketplaces.

if negative) and pays the dealer a fee $\phi_i(a) \in \mathbb{R}_+$ (in terms of the numeraire good) in exchange for the dealer's intermediation services. Both the quantity traded, $a_i^b$, and the fee, $\phi_i(a)$, will be determined through a bilateral bargaining procedure between the dealer and the investor.[14] We use $p \in \mathbb{R}_+$ to denote the price of the asset in the competitive market (also expressed in terms of the numeraire good). With instantaneous probability $\beta$ the investor gains direct access to the competitive asset market, and he simply chooses a new portfolio $a_i^c$ at cost $p(a_i^c - a)$.

The value function of a dealer, $W$, satisfies

$$rW = \alpha \int \phi_i(a) dH(a,i), \tag{2}$$

where $H(a,i)$ is the distribution of portfolios and preference types across investors. A dealer meets an investor with instantaneous probability $\alpha$. Random matching implies that this investor is a random draw from the population of all investors. Thus, (2) simply equates the flow value of a dealer to the expected intermediation fee.

We now turn to the determination of the intermediation fee and the quantity of assets traded in a meeting between a dealer and an investor of type $i$ who holds portfolio $a$. The change in the investor's portfolio, $a_i^b - a$, and the payment to the dealer, $\phi_i(a)$, are taken to be the outcome corresponding to the Nash solution to a bargaining problem where the dealer has bargaining power $\eta \in [0,1]$. The utility of an investor if an agreement $(a^b - a, \phi)$ is reached is $V_i(a^b) - p(a^b - a) - \phi$. In case of disagreement, the utility of the investor is $V_i(a)$. Therefore, the investor's surplus is $V_i(a^b) - V_i(a) - p(a^b - a) - \phi$. The dealer's surplus is equal to the fee, $\phi$. Hence, the outcome of the bargaining is given by[15]

$$(a_i^b, \phi_i) = \arg\max_{(a',\phi)} [V_i(a') - V_i(a) - p(a' - a) - \phi]^{1-\eta} \phi^{\eta}. \tag{3}$$

The following lemma characterizes the bargaining solution given the value functions $\{V_i\}_{i=1}^{I}$.

[14] Our notation reflects the fact that both the new portfolio and the intermediation fee may depend on the preference type of the investor at the time he contacts the dealer. In principle, these variables may also depend on the investor's wealth, but anticipating Lemma 1, we do not make this dependence explicit in our notation for the new portfolio.

[15] Note that it would be equivalent to set $\phi = (\hat{p} - p)(a^b - a)$ and reformulate the bargaining problem as a choice of $(a^b - a)$, the size of the order, and $\hat{p}$, the transaction price charged or paid by the dealer. So if $a^b > a$, then the investor is a buyer and $\hat{p} > p$ can be interpreted as the *ask price* charged by the dealer. Conversely, if $a^b < a$, then the investor is a seller and $\hat{p} < p$ is the *bid price* paid by the dealer.

**Lemma 1** *The outcome of the bargaining problem (3) is* $(a_i^b, \phi_i) = [a_i, \phi_i(a)]$*, where*

$$a_i = \arg\max_{a'}[V_i(a') - pa'], \tag{4}$$

$$\phi_i(a) = \eta \max_{a'}[V_i(a') - V_i(a) - p(a' - a)]. \tag{5}$$

According to Lemma 1, the quantity of assets the investor buys, $a_i - a$, is chosen so as to maximize the total surplus of a match and the intermediation fee, $\phi_i(a)$, is set to split the surplus of the match according to each agent's bargaining power. From (4), it is immediate that the investor's new portfolio, $a_i$, is independent of $a$. Interestingly, from the last term of (1), it is clear that investors choose the same portfolio, $a_i$, regardless of whether their trade is carried out directly in the asset market or intermediated by a dealer. We can use Lemma 1 to rewrite (1) as

$$rV_i(a) = u_i(a) + \delta \sum_j \pi_j [V_j(a) - V_i(a)] + \kappa \max_{a'}[V_i(a') - V_i(a) - p(a' - a)], \tag{6}$$

where $\kappa \equiv \alpha(1-\eta) + \beta$ can be thought of as the rate at which the investor gains *effective* direct access to the asset market.

So far we have characterized the bargaining outcome for given value functions and provided the maximizers corresponding to the right-hand sides of (4) and (5) exist. The following lemma establishes that the value functions exist, are unique, strictly increasing and strictly concave, that the $a_i$ defined in (4) is unique and that the $\phi_i(a)$ given in (5) is well-defined. For the analysis that follows, it will be convenient to define

$$U_i(a) = \frac{r+\kappa}{r+\delta+\kappa} u_i(a) + \frac{\delta}{r+\delta+\kappa} \sum_j \pi_j u_j(a). \tag{7}$$

**Lemma 2** *For each* $i$*, suppose that* $u_i(a) - rpa$ *is continuous and bounded above for any* $rp > 0$*. There exists a unique solution* $\{V_i(\cdot)\}_{i=1}^I$ *to (6):*

$$V_i(a) = \frac{U_i(a) + \kappa(pa + \Omega_i)}{r+\kappa}, \tag{8}$$

*where* $\Omega_i = \frac{r+\kappa}{r+\delta+\kappa}\Delta_i + \frac{\delta}{r+\delta+\kappa}\sum_j \pi_j \Delta_j$*, and* $\Delta_i = \max_x [U_i(x) - rpx]$*.*

Combining (4) and (8) we find that $a_i$, the optimal portfolio of an investor with preference type $i$, is the one that achieves the value $\Delta_i$ in Lemma 2. Hence, $a_i$ satisfies

$$U_i'(a_i) \le rp, \qquad \text{“}=\text{”} \quad \text{if } a_i > 0. \tag{9}$$

Condition (9) states that an investor who wishes to hold the asset in state $i$ chooses his optimal portfolio so that the expected (conditional on being in state $i$) discounted sum of flow marginal utility from holding the portfolio until the next time it is adjusted, i.e., $\frac{U_i'(a)}{r+\kappa}$, equals $p - \frac{\kappa}{r+\kappa}p$, namely the cost of buying an additional unit of the asset and holding it until the next time it can be sold back in the competitive market.

From (7) we see that $U_i'(a)$ is a weighted average of the marginal utilities in the various states. The weights on the current marginal utility and future ones depend on the transition rates $\alpha$, $\beta$ and $\delta$, on the discount rate $r$, on the dealer's bargaining power, $\eta$, and on the probability distribution $\{\pi_k\}_{k=1}^I$. If $u_i'(a_i) > \sum_k \pi_k u_k'(a_i)$, then $\partial a_i/\partial\alpha > 0$ and $\partial a_i/\partial\beta > 0$. The investor's demand for the asset increases as trading frictions are reduced. As $\alpha$ or $\beta$ increases, it becomes easier for investors to sell parts of their portfolios in case of an adverse preference shock. Therefore, the investor's demand for the asset increases in those states in which his marginal utility from holding the asset is high relative to the average across states. Conversely, if $u_i'(a_i) < \sum_k \pi_k u_k'(a_i)$, i.e., if his marginal utility in the current state is below average, then $\partial a_i/\partial\alpha < 0$ and $\partial a_i/\partial\beta < 0$. As $\alpha$ or $\beta$ goes to infinity, $U_i(a) \to u_i(a)$, and the optimal portfolio tends to the $a_i$ that solves $u_i'(a_i) \leq rp$, namely the portfolio choice that would prevail in a competitive market where all trades can be executed instantaneously.[16]

We now turn to the determination of the intermediation fee that an agent in state $i$ who is holding portfolio $a$ pays the dealer who readjusts his portfolio. From (5), this fee satisfies $\phi_i(a) = \eta\left[V_i(a_i) - V_i(a) - p(a_i - a)\right]$, with $a_i$ characterized by (9). Using (8),

$$\phi_i(a) = \frac{\eta\left[U_i(a_i) - U_i(a) - rp(a_i - a)\right]}{r+\kappa}. \tag{10}$$

The intermediation fee depends on the dealer's bargaining power, $\eta$, the discount factor, $r$, and the transition rates $\alpha$, $\beta$ and $\delta$. It also varies with the investor's asset position at the time the trade is executed, $a$, as well as with his desired portfolio, $a_i$.

Next, we characterize the steady-state distribution $H(a,i)$. The individual state of an investor is a pair $(a,i) \in \mathbb{R}_+ \times \{0, ..., I\}$, where $a$ is his current portfolio and $i$ his preference type. First, note that any state $(a,i)$ such that $a \neq a_j$ for $j \in \{1, ..., I\}$ is transient, since whenever an investor adjusts his portfolio he chooses $a \in \{a_k\}_{k=1}^I$. Thus, the set of ergodic

[16] Note that for $\alpha < \infty$ and $\beta < \infty$, the asset would appear to be misallocated in the sense that $u_i'(a_i) \neq u_j'(a_j)$ for $i \neq j$. Specifically, investors with relatively low current marginal valuations, i.e., those with preference types $i$ such that $u_i'(a_i) < \sum_k \pi_k u_k'(a_i)$, hold positions that are too large relative to their optimal portfolios in the frictionless benchmark. And similarly, investors with high current marginal valuations hold positions that are too small.

states is $\{a_k\}_{k=1}^I \times \{1, ..., I\}$. This allows us to simplify the exposition by denoting state $(a_i, j) \in \{a_k\}_{k=1}^I \times \{1, ..., I\}$ by $ij \in \{1, ..., I\}^2$. Hence, for state $ij$, $i$ represents the portfolio the investor currently has (i.e., the one corresponding to the preference shock he had at the time he last rebalanced his portfolio), and $j$ represents his current preference shock. The measure of investors in state $ij$ is denoted $n_{ij}$.

In a steady state, the flow of investors entering state $ij$ must equal the flow of investors leaving state $ij$:

$$\delta\pi_j \sum_{k\neq j} n_{ik} - \delta(1-\pi_j)n_{ij} - (\alpha+\beta)\, n_{ij} = 0, \quad \text{if } j \neq i, \tag{11}$$

$$(\alpha+\beta) \sum_{k\neq i} n_{ki} + \delta\pi_i \sum_{k\neq i} n_{ik} - \delta(1-\pi_i)n_{ii} = 0. \tag{12}$$

According to (11), the measure of investors in state $ij$ ($j \neq i$) increases whenever an investor in some state $ik$ receives a preference shock $j$ (which occurs with instantaneous probability $\delta\pi_j$) and decreases whenever an investor in state $ij$ receives a new preference shock different from $j$ (which happens with instantaneous probability $\delta(1-\pi_j)$), or whenever such an investor readjusts his portfolio (either through a dealer, with instantaneous probability $\alpha$, or by accessing the asset market directly, with instantaneous probability $\beta$). Equation (12) has a similar interpretation. The pattern of flows between states is depicted in Figure 2 for an example with $I = 3$. Each circle represents an individual state. The horizontal arrows represent flows due to preference shocks, whereas the vertical ones indicate flows due to portfolio readjustments. On the diagonal, the individual states shaded in grey are those for which there is no mismatch between the investor's current preference type and the portfolio he holds. The following lemma characterizes the stationary distribution of preference shocks and asset holdings.

**Lemma 3** *The steady-state distribution* $(n_{ij})_{i,j=1}^I$ *is given by*

$$n_{ij} = \frac{\delta\pi_i\pi_j}{\alpha+\beta+\delta}, \quad \textit{for } j \neq i, \tag{13}$$

$$n_{ii} = \frac{\delta\pi_i^2 + (\alpha+\beta)\pi_i}{\alpha+\beta+\delta}. \tag{14}$$

The marginal distributions $n_{i\cdot} = \sum_j n_{ij}$ and $n_{\cdot j} = \sum_i n_{ij}$ are $n_{i\cdot} = n_{\cdot i} = \pi_i$, i.e., the measure of investors with preference type $i$ is equal to $\pi_i$, the probability to draw preference shock $i$, conditional on getting a preference shock. Note that the distribution of probabilities

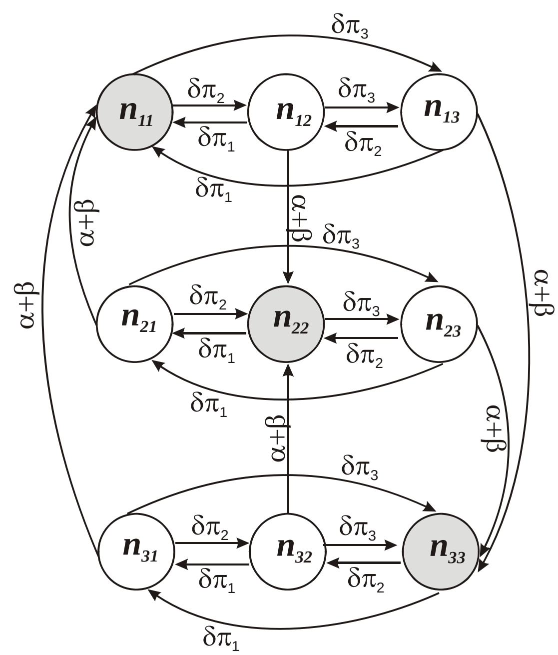


Figure 2: Flows between individual investor states

across states is symmetric, $n_{ij} = n_{ji}$. Also, $\partial n_{ij}/\partial(\alpha+\beta) < 0$ if $j \neq i$ and $\partial n_{ii}/\partial(\alpha+\beta) > 0$, i.e., the measure of investors who are matched to their desired portfolio increases as the rate at which investors get to rebalance their portfolios, either directly or through dealers, increases.

So far we have characterized the optimal portfolio choice, $a_i$, for an investor of type $i$, with asset position $a$, the fee $\phi_i(a)$ that the investor pays if the trade is intermediated by a dealer and the steady-state distribution of investors across asset holdings and preference types. The only remaining equilibrium variable to be determined is $p$, the price of the asset in the competitive market. This price is the one that equilibrates demand and supply of assets, i.e., the one that implies $\sum_{i,j} n_{ij} a_i = A$. Using Lemma 3, this market-clearing condition can be written as

$$\sum_i \pi_i a_i = A. \tag{15}$$

A steady-state equilibrium is a list $\{(n_{ij})_{i,j=1}^{I}, (a_i, \phi_i(\cdot))_{i=1}^{I}, p\}$ such that $(n_{ij})_{i,j=1}^{I}$ satisfies (13) and (14), $(a_i,)_{i=1}^{I}$ and $p$ satisfy (9) and (15), and $(\phi_i(\cdot))_{i=1}^{I}$ satisfies (10).

**Proposition 1** *There exists a unique steady-state equilibrium.*

The distribution of investors over portfolios and preference types is given by (13) and (14). The individual portfolio choices ($a_i$) in (9) depend on $p$, the equilibrium price in the interdealer market. Given these individual demands, the market-clearing condition (15) determines a unique price.

Hereafter, unless it is specified otherwise, suppose $u_i$ is such that (9) holds at equality (e.g., suppose that $\lim_{a\to 0} u_i'(a) = \infty$ for some $i$). To illustrate how a reduction in trading frictions affects the equilibrium, we consider first the limiting case where search frictions vanish, i.e., as either $\alpha \to \infty$ or $\beta \to \infty$. In either case investors can trade in the asset market continuously, as they can either find dealers instantly (if $\alpha = \infty$) or have continuous direct access to the asset market (if $\beta = \infty$). In the limit, from (9) we get

$$\frac{u_i'(a_i)}{r} = p \tag{16}$$

for $i = 1, ..., I$. Combining (15) and (16), we see that the price of the asset converges to the $p$ that solves $\sum_i \pi_i u_i'^{-1}(rp) = A$. From (10) we see that $\phi_i(a) \to 0$ for all $a$ and $i$. The limiting distribution of investors across asset holdings and preference types is $n_{ii} = \pi_i$ for each $i$, and $n_{ij} = 0$ for $j \neq i$. As frictions vanish, investors choose $a_i$ continuously by equating the present discounted value of the marginal return from the asset to its price. The equilibrium fee, asset price and distribution of asset positions are the ones that would prevail in a Walrasian economy.[17]

Another interesting limiting case results when $\eta = 1$ (dealers have all the bargaining power). In this special case, the equilibrium is still characterized by (9), (10), (13), (14), and (15), but with

$$U_i(a) = \frac{r+\beta}{r+\beta+\delta} u_i(a) + \frac{\delta}{r+\beta+\delta} \sum_k \pi_k u_k(a).$$

Since investors enjoy no gains from readjusting their portfolios in trades intermediated by dealers, the demand for assets is the one that would prevail in an economy with $\alpha = 0$. The liquidity provided by dealers is irrelevant for investors' portfolio choices, and investors behave as if there were no dealers to alleviate the trading frictions.

[17]For related limiting results, see Duffie, Gârleanu and Pedersen (2005) and Miao (2006). In a different context, see Gale (1987) and Spulber (1996).

### 3.1 Aggregate uncertainty

We now extend the model to allow for aggregate uncertainty. Suppose the economy can be in one of two states: a high state ($H$) in which every investor of preference type $i$ enjoys utility $u_i^H(a)$ from holding portfolio $a$ and a low state ($L$) in which every investor of preference type $i$ enjoys utility $u_i^L(a)$ from the same portfolio, with $\partial u_i^L(a)/\partial a < \partial u_i^H(a)/\partial a$. The aggregate state changes according to a continuous-time Markov chain: while in state $s \in \{H, L\}$, the time until the next switch to the other state $s'$ is an exponentially distributed random variable with mean $1/\lambda^s$. So, for example, if we associate state $H$ with "normal times" and state $L$ with "times of crisis," a small $\lambda^H$ means that crises are very infrequent, and a large $\lambda^L$ that they tend to be short-lived. We specialize the analysis to stochastic steady states where $n_i(t) = \pi_i$ for all $t$. Let $p^s$ denote the price of the asset when the aggregate state is $s$.

Let $V_i^s(a)$ denote the value of an investor with preference type $i$ when the aggregate state is $s \in \{H, L\}$. It satisfies a generalized version of (6), i.e.,

$$\begin{aligned} rV_i^s(a) = u_i^s(a) + \delta \sum_j \pi_j [V_j^s(a) - V_i^s(a)] + \lambda^s [V_i^{s'}(a) - V_i^s(a)] \\ + \kappa \max_{a'} \{[V_i^s(a') - V_i^s(a) - p^s(a' - a)]\}, \qquad (17) \end{aligned}$$

for $i = 1, ..., I$, $s = H, L$ and $s' \in \{H, L\} \setminus \{s\}$. The novelty in (17) is the third term on the right side, which captures the capital gain associated with a change in the aggregate state. The investor's problem can be characterized following a similar method to the one we used in Section 3. For the analysis that follows, it will be convenient to define $\lambda \equiv \lambda^H + \lambda^L$ and

$$U_i^s(a) = \frac{(r+\kappa)[u_i^s(a) + v_i^s(a)] + \delta \sum_j \pi_j [u_j^s(a) + v_j^s(a)]}{r + \delta + \kappa},$$

with $v_i^s(a) = \frac{(r+\kappa+\lambda)[u_i^{s'}(a) - u_i^s(a)] + \delta \sum_j \pi_j [u_j^{s'}(a) - u_j^s(a)]}{(1/\lambda^s)(r+\kappa+\lambda)(r+\kappa+\delta+\lambda)}$, for $s = H, L$ and $s' \in \{H, L\} \setminus \{s\}$. The following lemma summarizes the portfolio problem that investors face in this environment.

**Lemma 4** *An investor with preference type $i \in \{1, ..., I\}$ who gains access to the asset market at a time when the aggregate state is $s \in \{H, L\}$ solves*

$$\max_{a_i^s} [U_i^s(a_i^s) - \xi^s a_i^s], \qquad (18)$$

*where*

$$\xi^s = rp^s - \frac{\kappa \lambda^s}{r + \lambda + \kappa}(p^{s'} - p^s), \qquad (19)$$

*for $s = H, L$ and $s' \in \{H, L\} \setminus \{s\}$.*

Intuitively, $\frac{U_i^s(a_i^s)}{r+\kappa}$ is the present expected discounted utility that the investor obtains from holding $a_i^s$ over the period of time that elapses until he is next able to rebalance his portfolio, while $\frac{\xi^s}{r+\kappa}$ is the cost of purchasing a unit of the asset today, net of its expected discounted resale value at the next time he regains access to the market.[18] Thus, the portfolio choice of an investor of type $i$ in state $s$ satisfies

$$\frac{\partial U_i^s(a_i^s)}{\partial a_i^s} \leq \xi^s, \qquad \text{“}=\text{”} \quad \text{if } a_i^s > 0. \tag{20}$$

In order to characterize the price of the asset in each state, we turn to the market-clearing condition. The flow of assets brought by investors who access the market, directly or through a dealer, is $(\alpha + \beta) \int a dH_t(a, i) = (\alpha + \beta) A$. Among those investors, a fraction $\pi_i$ of them have preference type $i$. Therefore, the aggregate demand for the asset is $(\alpha + \beta) \sum_i \pi_i a_i^s$ and the market-clearing conditions can be written as

$$\sum_i \pi_i a_i^s = A, \text{ for } s = H, L. \tag{21}$$

Since (20) gives $a_i^s$ as a function of $\xi^s$, (21) can be solved for the pair $(\xi^H, \xi^L)$, and given this pair, asset prices can be obtained from (19), i.e.,

$$p^s = \frac{\xi^s}{r} - \frac{\kappa \lambda^s}{r\,(r+\kappa)\,(r+\lambda)} (\xi^s - \xi^{s'}), \tag{22}$$

for $s = H, L$ and $s' \in \{H, L\} \setminus \{s\}$. The intermediation fee borne by an investor of preference type $i$ who purchases $(a_i^s - a)$ through a dealer at a time when the aggregate state is $s$ satisfies a generalized version of (5), $\phi_i^s(a) = \eta[V_i^s\,(a_i^s) - V_i^s(a) - p^s\,(a_i^s - a)]$. We can substitute $V_i^s(\cdot)$ (e.g., using (52) in the appendix) to arrive at

$$\phi_i^s(a) = \frac{\eta\,[U_i^s\,(a_i^s) - U_i^s\,(a) - \xi^s\,(a_i^s - a)]}{r+\kappa}. \tag{23}$$

A stationary stochastic equilibrium is a list $\{(n_i)_{i=1}^I, ((a_i^s, \phi_i^s\,(\cdot))_{i=1}^I, \xi^s, p^s)_{s=H,L}\}$ such that $(n_i)_{i=1}^I = (\pi_i)_{i=1}^I$; given $(\xi^s)_{s=H,L}$, $a_i^s$ satisfies (20) for all $i$, $s$; $(\xi^s)_{s=H,L}$ satisfy (21); $(p^s)_{s=H,L}$ are given by (22); and $\phi_i^s\,(\cdot)$ satisfies (23) for all $i$ and $s$. It is possible to show that a stationary stochastic equilibrium exists and is unique. For each aggregate state $s$, (20) and (21) determine $(a_i^s)_{i=1}^I$ and $\xi^s$. Given the pair $(\xi^H, \xi^L)$, (22) implies the equilibrium prices $(p^H, p^L)$. Given $((a_i^s)_{i=1}^I, p^s)_{s=H,L}$, the expression for the intermediation fee is immediate from (23).

[18] Note that $\frac{\xi^s}{r+\kappa} = p^s - \mathbb{E}_s[e^{-rT} p^{w(T)}]$, where the expectation is over the random variables $T$ and $w\,(T) \in \{H, L\}$, namely the length of time until the date at which the investor will regain access to the market and the aggregate state that will prevail at that date. (The subscript indicates that the expectation is conditional on the current aggregate state, $s$; see the appendix for details.)

### 3.2 Numerical analysis

In this section we parametrize the model and use it, here and in subsequent sections, to illustrate and complement our analytical results. We let a unit of time correspond to a day and normalize the stock of assets, $A = 1$. We take the rate of time preference to be 7 percent per year, i.e., $r = 1.07^{\frac{1}{360}} - 1$. As a benchmark, we consider a pure dealer market where all trades are intermediated by dealers, i.e., we set $\beta = 0$ (we experiment with a wide range of values for $\beta$ in Sections 6 and 7). The average delay of execution for a trade intermediated by a dealer is taken to be one day, i.e., $\alpha = 1$.[19] We set $\delta = 1$ so that investors receive preference shocks at the same rate as they encounter trade opportunities. We assume the negotiating parties have equal bargaining power, i.e., $\eta = 0.5$. We let $u_i(a) = \varepsilon_i a^{1-\sigma}/(1-\sigma)$ and take $\sigma = 2$ as our baseline, but will report some results for various values of $\sigma$. The support for the values of $\varepsilon_i$ is $\{\frac{i-1}{I-1}\}_{i=1}^{I}$ with $I = 50$. Our choice of $I$ implies that investors will be distributed among 2,500 individual states. The preference shock $\varepsilon_i = \frac{i-1}{I-1}$ is drawn with probability

$$\pi_i = \frac{\lambda^{i-1}/(i-1)!}{\sum_{j=1}^{I} \lambda^{j-1}/(j-1)!}, \text{ for } i = 1, ..., I. \tag{24}$$

The parameter $\lambda$ from this truncated Poisson distribution plays a key role in determining the size distribution of trades and intermediation fees. We choose $\lambda = 4$. This parametrization implies a yearly turnover rate for the asset that is close to 8 and an average effective volume-weighted spread of about 0.2 basis points of the asset price.[20]

Figure 3 illustrates some key features of our economy. The distribution of idiosyncratic shocks is plotted on the top-left panel. The top-right panel shows investors' choices of portfolios as a function of the realization of their individual preference shocks. Not too surprisingly, $a_i$ increases with $\varepsilon_i$. The bottom-left panel is a three-dimensional histogram of the equilibrium stationary distribution of investors over individual states, i.e., $(n_{ij})_{i,j=1}^{I}$. The investor's desired

[19] According to Saunders, Srinivasan and Walter (2002, p. 97), trading delays in corporate bond markets range from a minute to a day.

[20] This value for the effective spread is smaller than estimates found in the literature—see Schultz (2001), for instance. One reason is that our model abstracts from other components of the bid-ask spread such as inventory costs and costs associated with asymmetric information between dealers and investors. The expression for the average bid-ask spread, weighted by volume and expressed as a proportion of the price of the asset, is given in (33). (See Section 6 for more on spreads.) The turnover in the model is defined as follows. The flow of investors who can readjust their portfolios per unit of time is $\alpha + \beta$. A fraction $n_{ji}$ of these investors readjust their portfolio from $a_j$ to $a_i$ so that the quantity they trade is $|a_i - a_j|$. Thus, the turnover rate is $\mathcal{T} = A^{-1} \sum_{i,j} (1/2)(\alpha + \beta) n_{ji} |a_i - a_j|$. Our turnover rate is in the same range as the turnover for the entire market for U.S. public and private debt. The outstanding debt is \$15.8 trillion while the daily trading volume is \$368 billion, so that the daily turnover rate is 0.023, or about 8 yearly (see Joys, 2001).

portfolio lies on the $x$ axis and his current portfolio on the $y$ axis. Both axes range from $a_1$ to $a_I$, which in this parametrization equal 0.82 and 2.19, respectively. There is a fairly large concentration of agents on the main diagonal, i.e., many investors hold their optimal portfolio. Visually, this is the spike that imitates the shape of the primitive distribution of idiosyncratic shocks. But note that there is also a significant proportion of agents experiencing various degrees of mismatch with their current portfolios, precisely 43 percent of them. The bottom-right panel plots the fees $\phi_{ij}$. Notice that fees are 0 on the main diagonal ($\phi_{ii} = 0$) and that they increase as $a_i$ (the current portfolio) and $a_j$ (the new portfolio) get further apart. Also, buying and selling fees are not symmetric.

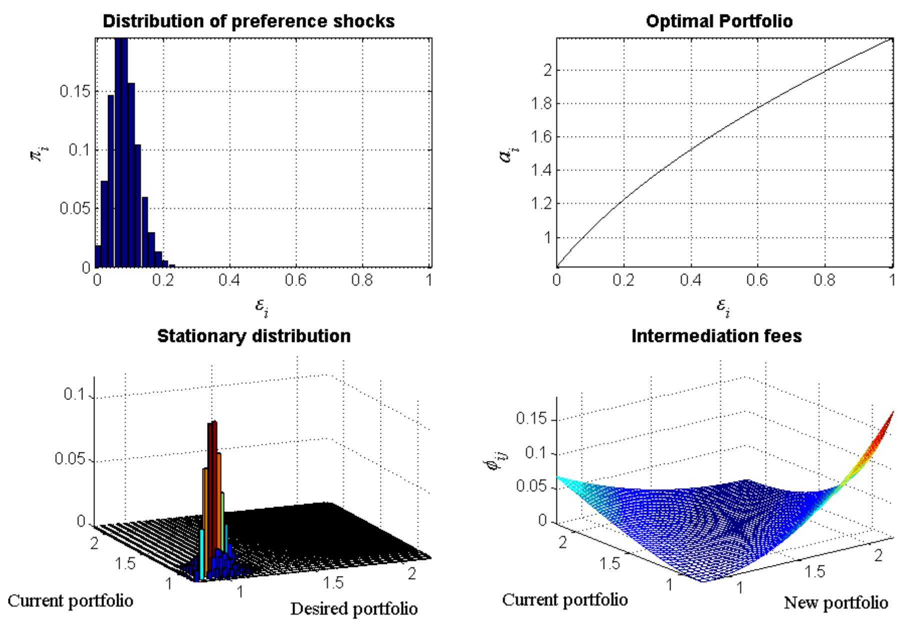


Figure 3: The baseline economy

# 4 Asset prices

In this section we discuss the effects of parameter changes on the level and volatility of asset prices. We will pay special attention to $\alpha$ and $\beta$, which we interpret as measures of execution delays. We specialize the analysis to utility functions of the form $u_i(a) = \varepsilon_i u(a)$. For this class of preferences, $U_i(a) = \bar{\varepsilon}_i u(a)$, where $\bar{\varepsilon}_i = \frac{(r+\kappa)\varepsilon_i + \delta\bar{\varepsilon}}{r+\delta+\kappa}$ and $\bar{\varepsilon} = \sum_j \pi_j \varepsilon_j$. Focusing on interior solutions, (9) reduces to

$$\bar{\varepsilon}_i u'(a) = rp. \tag{25}$$

### 4.1 Level

Differentiating (25) for a given $p$, we find that $\partial a_i/\partial\alpha$ has the same sign as $(\varepsilon_i - \bar{\varepsilon})$. That is, investors with a preference shock above average increase their demand when $\alpha$ increases. To see why this is so, note that the left side of (25) is a weighted sum of the investor's marginal utility from holding the asset. Agents with $\varepsilon_i > \bar{\varepsilon}$ have a current marginal utility that is higher than what they expect it to be in the future. Consequently, their choice of $a_i$ is lower than $u'^{-1}(rp/\varepsilon_i)$, which is what they would choose in a world with no trading delays. The reason is that, since $\varepsilon_i$ is higher than $\bar{\varepsilon}$, the investor anticipates that his preferences are likely to revert toward $\bar{\varepsilon}$ in the future, and that when this happens, he may be unable to rebalance his portfolio for some time. A larger $\alpha$ means that it will be easier for the investor to find a dealer in the future, and this makes him put more weight on his current marginal utility from holding the asset relative to its expected value. Conversely, investors with a preference shock below average reduce their demand when $\alpha$ increases. So, for given $p$, as $\alpha$ increases the dispersion of asset holdings increases. The same logic applies to changes in $\beta$, and for similar reasons, $\partial a_i/\partial\delta$ has the same sign as $-(\varepsilon_i - \bar{\varepsilon})$.

Since for a given price $p$, the demands of investors with relatively low valuations ($\varepsilon_i < \bar{\varepsilon}$) fall, while those of investors with high valuations ($\varepsilon_i > \bar{\varepsilon}$) rise, the effect of an increase in $\alpha$ on the aggregate demand for assets—and therefore on the equilibrium price of the asset—is ambiguous in general. The following proposition provides sufficient conditions for the price of the asset to increase with $\alpha$.

**Proposition 2** *If $-[u'(a)]^2/u''(a)$ is strictly increasing in $a$, then $dp/d\alpha > 0$. If $-[u'(a)]^2/u''(a)$ is strictly decreasing in $a$, then $dp/d\alpha < 0$. If $-[u'(a)]^2/u''(a)$ is independent of $a$, then $dp/d\alpha = 0$.*

Whether an increase in $\alpha$ has a positive effect on the asset price depends on the curvature of the utility function. The reason is that this curvature determines the curvature of the individual demand for the asset as a function of $\bar{\varepsilon}_i$, i.e., $\partial a_i/\partial\bar{\varepsilon}_i = -[u'(a_i)]^2/[u''(a_i)rp]$. If $u(a) = \log a$ then $a_i$ is linear in $\bar{\varepsilon}_i$, and as one aggregates the individual changes in demands induced by an increase in $\alpha$, the increases in $a_i$ (for investors with values of $\varepsilon_i$ larger than $\bar{\varepsilon}$) and the decreases in $a_i$ (for investors with values of $\varepsilon_i$ lower than $\bar{\varepsilon}$) cancel each other out. As a result, $\alpha$ has no effect on the aggregate demand for assets nor on the equilibrium price.[21] If $u$ is not too concave, $a_i$ is a convex function of $\bar{\varepsilon}_i$. For this case, the increases in $a_i$ for relatively large values of $\varepsilon_i$ outweigh the decreases in $a_i$ for relatively low values of $\varepsilon_i$ and the aggregate demand for the asset increases in response to an increase in $\alpha$. In turn, this implies that the equilibrium price of the asset increases with $\alpha$. Conversely, the asset price is decreasing in $\alpha$ if $u$ is sufficiently concave.

Proposition 2 is illustrated in Figure 4, where we plot the individual demand for the asset, $a_i$, as a function of $\bar{\varepsilon}_i$ for the special case of $I = 2$. The figure is drawn under the assumption that $-[u'(a)]^2/u''(a)$ is strictly increasing in $a$, so that $a_i$ is a strictly convex function of $\bar{\varepsilon}_i$. Notice that $\pi_1\bar{\varepsilon}_1 + \pi_2\bar{\varepsilon}_2 = \bar{\varepsilon}$ and that $\pi_1 a_1 + \pi_2 a_2 = A$ in equilibrium. An increase in $\alpha$ corresponds to a mean-preserving increase in the spread of the distribution of $\bar{\varepsilon}_i$, i.e., $\bar{\varepsilon}_i$ shifts to $\bar{\varepsilon}_i'$, with $\bar{\varepsilon}_1' < \bar{\varepsilon}_1$, $\bar{\varepsilon}_2' > \bar{\varepsilon}_2$ and $\pi_1\bar{\varepsilon}_1' + \pi_2\bar{\varepsilon}_2' = \bar{\varepsilon}$. From Jensen's inequality, it is clear that $\pi_1 a_1' + \pi_2 a_2' > \pi_1 a_1 + \pi_2 a_2$ (aggregate demand increases with $\alpha$ for given $p$), so the asset price has to increase to clear the market.

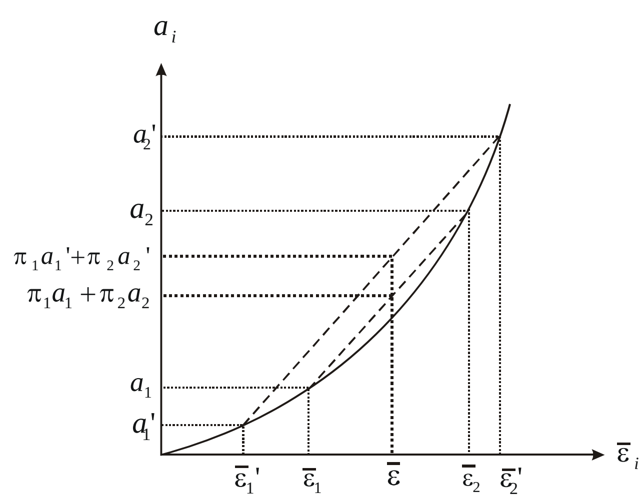


Figure 4: Effect of an increase in $\alpha$ on $p$

If we specialize preferences further by letting $u(a) = a^{1-\sigma}/(1-\sigma)$ with $\sigma > 0$, the model

[21] Gârleanu (2005) derives a similar result in a version of Duffie, Gârleanu and Pedersen (2005) with endogenous portfolios and CARA preferences.

can be solved in closed form:

$$a_i = \frac{\bar{\varepsilon}_i^{1/\sigma}}{\sum_j \pi_j \bar{\varepsilon}_j^{1/\sigma}} A \tag{26}$$

$$p = \frac{\left(\sum_i \pi_i \bar{\varepsilon}_i^{1/\sigma}\right)^{\sigma}}{r A^{\sigma}}. \tag{27}$$

As a corollary of Proposition 2, we know that $\sigma > 1$ ($< 1$) implies $dp/d\alpha < 0$ ($> 0$) and $u(a) = \log a$ implies $dp/d\alpha = 0$.

Finally, consider the limit as $\sigma \to 0$, i.e., as investors' preferences become linear. From (27), the price approaches $p = \bar{\varepsilon}_I / r$ and from (26) $a_i \to 0$ for $i \in \{1, ..., I-1\}$ and $a_I \to A/\pi_I$. The price is a weighted average of the marginal utility of the highest investor type and the average marginal utility in the market. The weight on the marginal utility of the highest valuation investor—and hence the asset price—is increasing in $\alpha$, $\beta$ and $r$, and decreasing in $\eta$ and $\delta$.

## 4.2 Volatility

We now investigate how trading frictions affect the volatility of asset prices. To this end, we consider the model with aggregate uncertainty developed in Section 3.1 and let $u_i^s(a) = z^s \varepsilon_i a^{1-\sigma}/(1-\sigma)$ where $z^s$ is the realization of the aggregate shock, with $z^H > z^L$. Then, $U_i^s(a) = \bar{\varepsilon}_i^s u(a)$, where $\bar{\varepsilon}_i^s$ is independent of $a$.[22] Consider an investor with idiosyncratic preference shock $i$ at a time when the aggregate state is $s$; his optimal portfolio is $a_i^s = (\bar{\varepsilon}_i^s / \xi^s)^{1/\sigma}$ and the market-clearing condition implies

$$\xi^s = \left[\frac{\sum_i \pi_i (\bar{\varepsilon}_i^s)^{1/\sigma}}{A}\right]^{\sigma}. \tag{28}$$

Consider first the limit that obtains as trading delays vanish. As $\kappa \to \infty$, we have $\bar{\varepsilon}_i^s \to z^s \varepsilon_i$, and combining (22) and (28) we find

$$p^s \to \frac{\left(\sum_i \pi_i {\varepsilon_i}^{1/\sigma}\right)^{\sigma}}{r A^{\sigma}} \frac{(r + \lambda^{s'}) z^s + \lambda^s z^{s'}}{r + \lambda}, \tag{29}$$

[22] Explicitly,

$$\bar{\varepsilon}_i^s = \frac{(r + \kappa + \lambda^{s'}) z^s + \lambda^s z^{s'}}{r + \kappa + \lambda} \bar{\varepsilon} + \frac{r + \kappa}{r + \kappa + \delta} \frac{(r + \kappa + \delta + \lambda^{s'}) z^s + \lambda^s z^{s'}}{r + \kappa + \delta + \lambda} (\varepsilon_i - \bar{\varepsilon}),$$

for $s = H, L$ and $s' \in \{H, L\} \backslash \{s\}$. Instead of considering aggregate preference shocks, we could assume that the asset yields a dividend stream $z$, where $z$ is a random variable. An investor holding portfolio $a$ would enjoy utility $\varepsilon_i u(za)$ while his idiosyncratic preference type is $i$.

for $s = H, L$ and $s' \in \{H, L\} \setminus \{s\}$.

Let us return to the case with finite $\kappa$ and set $\sigma = 1$. Using Lemma 4 and equations (22) and (28), we find $p^s = \frac{[(r+\lambda^{s'})z^s+\lambda^s z^{s'}]\bar{\varepsilon}}{r(r+\lambda)A}$. The price of the asset in state $s$ is independent of $\kappa$, the degree of trading frictions, and it is identical to the frictionless limit (29), with $\sigma = 1$. This finding is a generalization of the last part of Proposition 2 to economies with aggregate uncertainty: with logarithmic utility and multiplicative shocks, equilibrium prices are independent of trading frictions. Thus, in this case, trading frictions have no effect on the volatility of the asset price.

We use numerical simulations to study the effect that changes in the degree of trading frictions have on the volatility of the asset price for arbitrary values of $\sigma$. In a sufficiently long sample from our economy, the asset price has mean $\frac{\lambda^L p^H + \lambda^H p^L}{\lambda}$ and standard deviation $\frac{\sqrt{\lambda^H \lambda^L}}{\lambda}\left(p^H - p^L\right)$. From (29), in a frictionless economy the coefficient of variation (standard deviation divided by mean, hereafter CV), $\sqrt{\lambda^H \lambda^L}\left(p^H - p^L\right)/(\lambda^L p^H + \lambda^H p^L)$, equals

$$\frac{r\sqrt{\lambda^L \lambda^H}(z^H - z^L)}{(r+\lambda)\left(\lambda^L z^H + \lambda^H z^L\right)}.$$

According to this measure, is the volatility of the asset price higher in an economy with trading frictions?

To address this question we normalize $z^H = 1$, and merely for illustrative purposes, we let $z^L = 1/2$ and $\lambda^H = \lambda^L = 1/5$. The remaining parameter values are as in Section 3.2. For this parametrization, the ratio of the CV of the price in the economy with trading frictions to the CV in the frictionless economy is below 1, meaning that the asset price is more volatile in the frictionless economy than in the one with trading frictions. In fact, the asset price becomes more volatile as trading delays shrink. This result, however, depends crucially on the curvature of the utility function. To illustrate this fact, in Figure 5 we plot the level curves of the CV as a function of $\sigma$ and $\beta$ for the baseline parametrization of Section 3.2. This figure shows that the volatility of the asset price increases with $\beta$ if $\sigma > 1$, decreases with $\beta$ if $\sigma < 1$, and is independent of $\beta$ if $\sigma = 1$.[23]

[23] These results are mainly due to the fact that, in line with Proposition 2, the price of the asset in both states—and hence the denominator of the CV—is decreasing (increasing) in $\beta$ for $\sigma > 1$ ($< 1$).

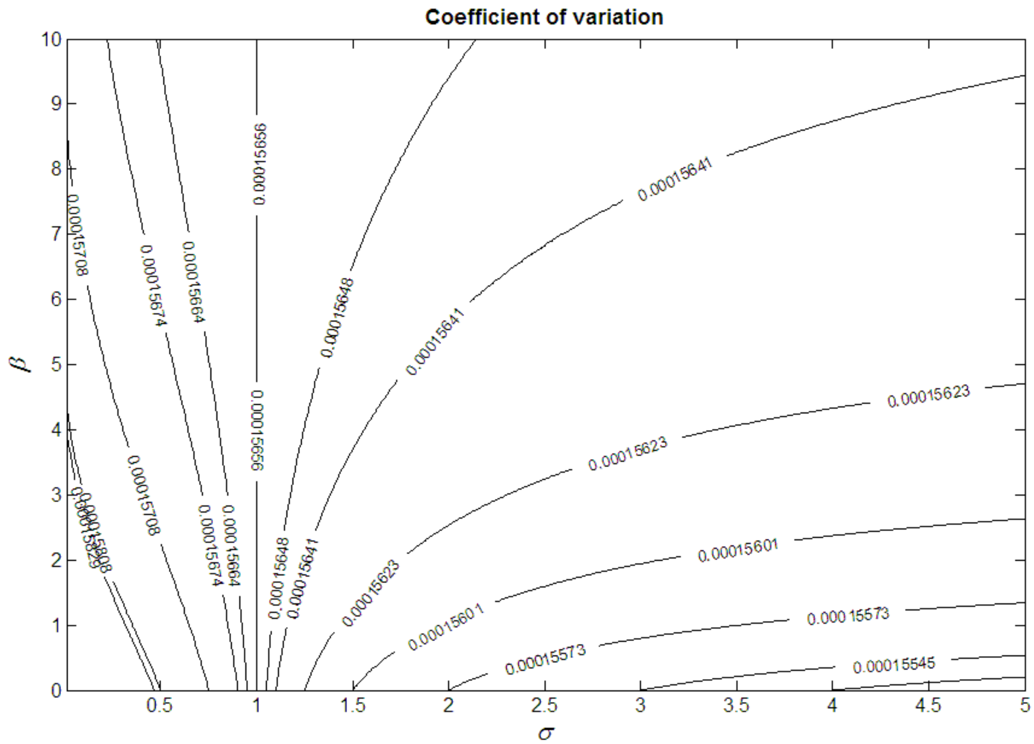


Figure 5: Asset-price volatility as a function of $\beta$ and $\sigma$

# 5 Trading frictions and volume

In this section we study the effects of trading frictions on the asset allocation across investors and the volume of trade. Let $\mathbb{G}_\kappa(a)$ denote the cumulative distribution of asset holdings across investors in an economy in which they gain direct effective access to the asset market at rate $\kappa$. For a particular class of utility functions, the following proposition establishes that the equilibrium distributions of asset holdings corresponding to different values of $\kappa$ can be ranked according to the second-order stochastic dominance ordering.

**Proposition 3** *Assume $u_i(a) = \varepsilon_i a^{1-\sigma}/(1-\sigma)$ with $\sigma > 0$. For any pair $(\kappa, \kappa')$ such that $\kappa' > \kappa$, $\mathbb{G}_\kappa$ dominates $\mathbb{G}_{\kappa'}$ in the second-order stochastic sense.*

Proposition 3 shows that the distribution of asset holdings across investors becomes "riskier," in a second-order stochastic sense, when trading frictions are reduced (or when investors have more bargaining power). The reason, as discussed in the previous sections, is that if an investor can access the market more frequently, he will choose a portfolio which is more in line with his current preference type. Hence, investors with a high $\varepsilon_i$ will raise $a_i$, while investors with a low $\varepsilon_i$ will reduce $a_i$. We now turn from the effects of trading frictions on the distribution of individual portfolios to their effects on the total volume of trade.

The flow of investors who can readjust their portfolios per unit of time is $\alpha+\beta$. A fraction $n_{ji}$ of these investors readjust their portfolio from $a_j$ to $a_i$ so that the quantity they trade is $|a_i - a_j|$. Thus, the total *volume of trade* is

$$\mathcal{V} = (\alpha+\beta)(1/2)\sum_{i,j} n_{ij}\left|a_j - a_i\right|. \tag{30}$$

(The *turnover rate* of the asset is defined to be $\mathcal{T} = \mathcal{V}/A$.) An increase in $\kappa$ has three distinct effects on trade volume. First, the measure of investors in any individual state $(i,j) \in I^2$ who gain access to the market and are therefore *able to trade* increases, which tends to increase trade volume. Second, the proportion $1-\sum_i n_{ii}$ of agents who are mismatched to their portfolio—and hence the fraction of agents who *wish to trade*—decreases, which tends to reduce trade volume. Finally, the distribution of asset holdings spreads out, which—according to Proposition 3—tends to increase the quantity of assets traded in many individual trades. With (13) and (30), it is easy to check that the first two effects combined lead to an increase in $\mathcal{V}$. In the case $I=2$, for example, it is also immediate from Proposition 3 that $|a_2 - a_1|$ increases with $\kappa$, so the total volume of trade unambiguously increases with $\kappa$. More generally, in our parametrization of Section 3.2 with $I = 50$, we also find $\partial\mathcal{V}/\partial\kappa > 0$. The behavior of the (daily) trade volume, $\mathcal{V}$, as a function of $\beta$ is illustrated in the bottom-right panel of Figure 6.

# 6 Trading frictions and spreads

In this section we study how changes in trading frictions affect intermediation fees, which we can also interpret as bid-ask spreads.[24] We specialize the analysis to $u_i(a) = \varepsilon_i a^{1-\sigma}/(1-\sigma)$, for $\sigma > 0$. From (10), the equilibrium fee that a dealer charges an investor who holds portfolio $a$ and wishes to hold $a_i$ is

$$\phi_i(a) = \frac{\eta\left[\frac{\bar{\varepsilon}_i}{1-\sigma}\left(a_i^{1-\sigma} - a^{1-\sigma}\right) - rp\left(a_i - a\right)\right]}{r+\kappa}, \tag{31}$$

[24] Recall the discussion in footnote 15.

with $a_i$ and $p$ given by (26) and (27). From (31), we see that an increase in $\beta$ (or $\alpha$) has two opposite effects on the intermediation fee. On the one hand, a higher $\beta$ implies more competition among dealers, which tends to reduce the fees they charge for any given trade size—the *competition effect* of reduced trading frictions. But on the other hand, a higher $\beta$ also induces investors to conduct larger portfolio reallocations every time they trade, and this translates into larger fees for dealers, on average—the *reallocation effect* of reduced trading frictions. Notice that along a stationary equilibrium the only transactions that investors carry out involve trading $a_j - a_i$, for $(i, j) \in I^2$. Therefore, we can simplify notation by letting $\phi_{ij}$ denote $\phi_j(a_i)$, namely the intermediation fee borne by an investor who holds portfolio $a_i$ and engages in a trade that leaves him with $a_j$.

Consider the limiting case $\sigma \to 0$, i.e., investors' preferences become almost linear. At the end of Section 4.1 we showed that in this case, $a_i \to 0$ for all $i \neq I$ and $rp \to \bar{\varepsilon}_I$, so (31) yields $\phi_{ij} \to 0$ for all $(i, j) \notin \{I\} \times \{1, ..., I-1\}$. Obviously, dealers obtain no fee when investors do not want to readjust their portfolios. Perhaps more surprisingly, when investors are buying the asset ($i \neq I$ and $j = I$), dealers do not charge a fee either. The reason is that when buying, the investor is paying his marginal valuation for the asset, and since he has almost-linear utility, this means that he is indifferent between holding or not holding the asset. Finally, there are investors in state $ij$, for $i = I$ and $j \neq I$. Those investors are holding $a_I \to A/\pi_I$ but wish to hold $a_j \to 0$. From (31), we find

$$\phi_{Ij} \to \frac{\eta(\varepsilon_I - \varepsilon_j) A}{(r + \delta + \kappa)\pi_I}, \tag{32}$$

a fee that is proportional to the quantity traded $(A/\pi_I)$.

Since the intermediation fee (32) is linear in the quantity traded, the previous results can be readily interpreted in terms of bid-ask spreads. The fact that an investor pays no fee when buying from the dealer is equivalent to a transaction in which the dealer charges an ask-price $p^a$ equal to the price of the asset in the competitive market, i.e., $p^a = p$. When an investor of type $j < I$ sells his portfolio $A/\pi_I$ through a dealer, he receives $pA/\pi_I - \phi_{Ij}$. Using (32), this transaction is equivalent to one in which the dealer pays investors of type $j$ a bid price $p_j^b = p - \frac{\eta(\varepsilon_I - \varepsilon_j)}{r+\delta+\kappa} < p$. The difference between the effective price at which the dealer sells, $p^a$, and buys, $p_j^b$, is akin to a bid-ask spread:

$$p^a - p_j^b = \frac{\eta(\varepsilon_I - \varepsilon_j)}{r + \delta + \kappa}.$$

This spread is decreasing in the rate of time preference (recall that agents only get returns from selling in this linear case) and in the rate at which investors can rebalance their portfolios ($\kappa$). It also decreases with $\delta$, since the value of rebalancing the portfolio is lower when preference shocks are more frequent. The spread increases with the dealer's bargaining power $\eta$, and also with the difference between the marginal utility of the highest-valuation investor and the marginal valuation of the investor involved in the trade. (Dealers buy assets at a lower effective price from investors with low marginal utility of consumption because these investors experience larger gains from selling the portfolio.)

Although for more general (i.e., nonlinear) preferences intermediation fees are nonlinear in the quantities traded, one can still compute the *effective prices* that an investor pays (or receives) per unit of the asset he buys (or sells). For example, investors with asset position $a_i$ who trade quantity $a_j - a_i$ through a dealer pay (or receive if $a_j - a_i$ is negative)

$$\hat{p}_{ij} = p + \frac{\phi_{ij}}{a_j - a_i}$$

per unit of the asset. The difference between the prices at which investors buy and sell is sometimes taken as a measure of the liquidity of the market.[25] Notice that if $a_j - a_i > 0$, then $\hat{p}_{ij} - \hat{p}_{ji} = \frac{\phi_{ij}+\phi_{ji}}{a_j - a_i} > 0$, so for this typical "round-trip" transaction, dealers (investors) trade at a lower (higher) effective price when they buy than when they sell. A financial analyst collecting transaction price data in this economy would compute an average *effective spread* weighted by volume and expressed as a proportion of the price of the asset as follows:[26]

$$\mathcal{S} = \frac{1}{p}\sum_{i,j} \frac{n_{ij}\,|a_i - a_j|}{\sum_{k,\ell} n_{k\ell}\,|a_k - a_\ell|}\frac{\phi_{ij}}{|a_i - a_j|}. \tag{33}$$

To illustrate the effects of trading frictions on intermediation fees, we turn to our baseline parametrization. The first panel of Figure 6 displays $\phi_{1I}$ and $\phi_{I1}$ as a proportion of the equilibrium asset price. The fee $\phi_{1I}$ paid by the investor who holds $a_1$ but wishes to trade up to $a_I$ (the top line) is hump-shaped with respect to $\beta$. For small $\beta$ the reallocation effect dominates and fees are increasing in $\beta$, but as $\beta$ gets larger, the competition effect between dealers begins

[25] See Harris (2003) for a textbook treatment, for example.

[26] Alternatively, one could compute the effective spread *weighted by the number of transactions*, namely

$$\sum_{j>i} \frac{n_{ij}}{1 - \sum_k n_{kk}} \frac{\phi_{ij} + \phi_{ji}}{a_j - a_i}.$$

to dominate, and the fee eventually becomes decreasing in $\beta$. In contrast, the fee $\phi_{I1}$ is strictly decreasing with $\beta$: the competition effect always outweighs the reallocation effect. Also, notice from the figure that buying and selling fees need not be symmetric, i.e., $\phi_{1I} \neq \phi_{I1}$.

A dealer's expected profit depends on the average fee he charges across the various trades. (This variable will play an important role when we endogenize the number of dealers in the following section.) The average fee, $\bar{\phi}$, equals $\sum_{i,j} \phi_{ij} n_{ij}$, or using (10),

$$\bar{\phi} = \eta \sum_{i,j} n_{ji} \frac{U_i(a_i) - U_i(a_j)}{r + \kappa}. \tag{34}$$

Thus, $\bar{\phi}$ depends on the mismatch between investors' desired and their actual portfolios, as measured by $U_i(a_i) - U_i(a_j)$, as well as on the frequency with which investors gain access to the asset market. More asset mismatch implies a larger expected return from providing intermediation services. An increase in the frequency at which an investor can have direct access to the market has several effects on dealers' profits. First, it raises investors' outside option and therefore reduces the fees dealers can charge in any given transaction. Second, it reduces the measure of investors who do not hold their desired portfolios (i.e., $n_{ii}$ increases for all $i$). Third, it raises the dispersion of portfolios and therefore the fees that dealers can charge to allow an investor to rebalance his portfolio. As a result, the average fee can be non-monotonic as illustrated in the top-right panel of Figure 6, which reports $\bar{\phi}$ (but normalized by the price of the asset) as a function of $\beta$. For low values of $\beta$ the dealers' expected profit increases because the dispersion of portfolios increases as trading frictions are reduced. But for sufficiently large values of $\beta$ the competition effect dominates and the average remuneration of dealers falls. The bottom-left panel of Figure 6 displays the average spread in (33), which is decreasing in $\beta$. Finally, the bottom-right panel shows that trade volume as defined in (30)—another variable that practitioners routinely associate with the "liquidity of the market"—is increasing in $\beta$.

The theory also has clear predictions for how individual intermediation fees vary with the size of a transaction. We summarize these implications in the following proposition.

**Proposition 4** *Consider a trade between a dealer and an investor holding portfolio $a$ and wishing to hold portfolio $a_i$. The intermediation fee per unit of asset traded paid by this investor, i.e., $\left|\frac{\phi_i(a)}{a_i - a}\right|$, is increasing in the size of the trade.*

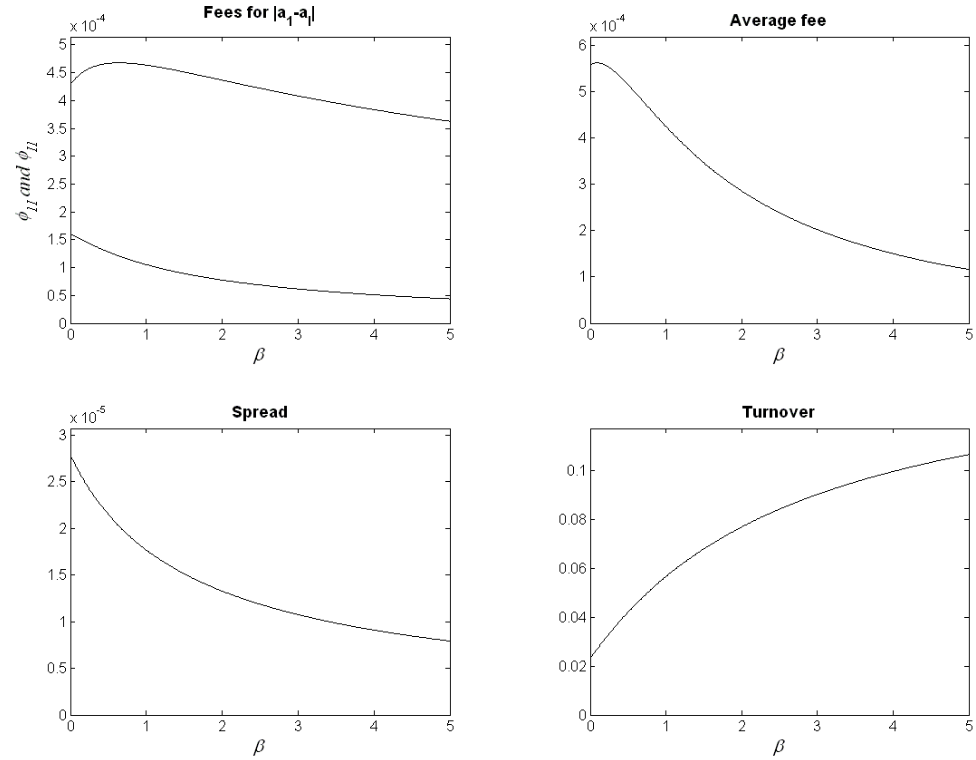


Figure 6: Trading frictions, intermediation costs and turnover

Figure 7 displays the fees per unit of asset traded for all the transactions that take place in the baseline parametrization of Section 3.2. In other words, it plots all the pairs $(a_j - a_i, \phi_{ij})_{j=1}^{I}$ for each $i = 1, ..., I$. As a corollary of Proposition 4, the intermediation fee per unit of asset traded tends to increase with the size of the trade.[27] Note, however, that two trades of the same size can pay different per-unit fees, since the associated surpluses for the investors may be different in the two trades. That is, in general, if $a_i \neq a_k$, then $\phi_{ij} \neq \phi_{ks}$ even if $a_j - a_i = a_s - a_k$. There is also an interesting asymmetry in terms of fees that are charged when investors buy vis-à-vis those they are charged when they sell. For example, an agent who buys $a_I - a_1$ pays a fee that is more than 2.5 times higher than the one the same dealer would charge to an investor who *sells* $a_I - a_1$. Our analytical insights—e.g., that with linear preferences only sellers pay

[27] This result is in accordance with Boehmer (2005, Table 7, Panel B), who finds that the effective spread increases with the size of orders in U.S. equity markets. In contrast, Edwards, Harris and Piwowar (2004) argue that transaction costs decrease significantly with trade size in corporate bond markets.

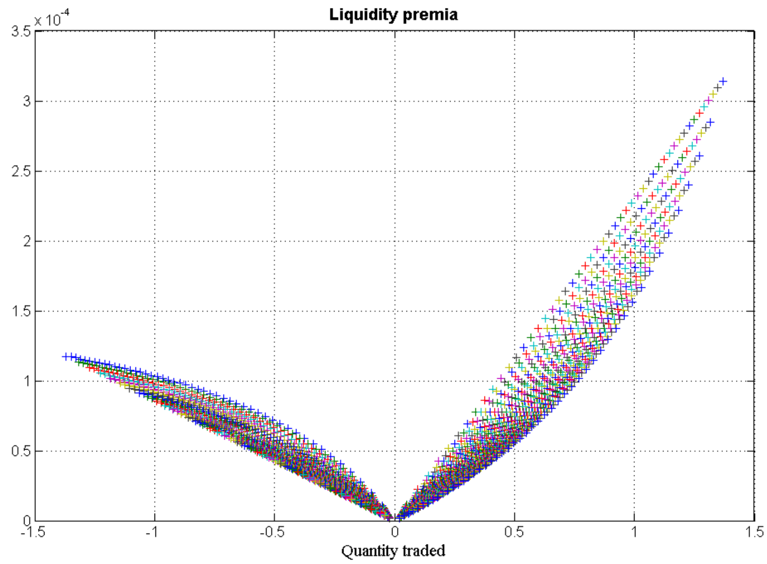


Figure 7: Liquidity premia

fees—suggest that this asymmetry is intimately linked to the value of $\sigma$. Our quantitative work confirms this conjecture: keeping the size of the trade constant, for larger values of $\sigma$, buyers pay larger fees than sellers.[28]

# 7 Entry of dealers

> In competitive dealer markets, dealer spreads ultimately depend on the costs that dealers incur in running their business. The free entry and exit of dealers ensures that spreads will adjust so that dealers just earn normal profits. When spreads are too high, their competition for order flow will cause spreads to fall, and as spreads fall, so do expected profits. Harris (2003, p. 298)

In this section we formalize the notion that a dealer's expected profits depend on the competition for order flow that he faces from other dealers. Many dealer markets are characterized by a virtual absence of barriers to entry.[29] Accordingly, we extend the model to allow for free entry

[28] In terms of this asymmetry, there is nothing special about $\sigma = 1$. In fact, buyers pay higher fees than sellers when $\sigma = 1$ in our baseline parametrization.

[29] See Wahal (1997) or Weston (2000) for an empirical study of the determinants of entry and exit of market-makers in NASDAQ and their impact on spreads and the level of trading activity, e.g., trade volume and number

of dealers, thereby endogenizing the speed at which investors can rebalance their portfolios.

Suppose that the Poisson rate at which an investor contacts a dealer, $\alpha$, is a continuously differentiable function of the measure of dealers in the market, $\upsilon$, with $\alpha(\upsilon)$ a strictly increasing and $\alpha(\upsilon)/\upsilon$ a strictly decreasing function of $\upsilon$. We specify that $\alpha(0) = 0$, $\alpha(\infty) = \infty$ and $\alpha(\infty)/\infty = 0$. Since all matches are bilateral and random, the Poisson rate at which a dealer serves an investor is $\alpha(\upsilon)/\upsilon$. For larger $\upsilon$, investors' orders are executed faster, but the flow of orders per dealer decreases due to a congestion effect.

There is a large measure of dealers who can choose to participate in the market. Dealers who choose to operate incur a flow cost $\gamma > 0$ that represents the ongoing costs of running the dealership, e.g., exchange membership dues, the cost of searching for investors, advertising their services and so on. Free entry implies $\frac{\alpha(\upsilon)}{\upsilon}\bar{\phi} = \gamma$, i.e., that the expected instantaneous profit of a dealer equals his flow operation cost.[30] Using (34), this condition can be rewritten as

$$\frac{\alpha(\upsilon)}{\upsilon}\eta\sum_{i,j} n_{ji}\frac{U_i(a_i) - U_i(a_j)}{r + \beta + \alpha(\upsilon)(1-\eta)} = \gamma. \tag{35}$$

A steady-state equilibrium with free entry is a list $\{(n_{ij})_{i,j=1}^{I}, (a_i, \phi_i(\cdot))_{i=1}^{I}, p, \upsilon\}$ that satisfies (9), (10), (13), (14), (15), and (35).

**Proposition 5** *Assume $\eta > 0$. There exists a steady-state equilibrium with free entry of dealers, and it has $\upsilon > 0$.*

Proposition 5 establishes the existence of a steady-state equilibrium with free entry provided dealers have some bargaining power. (If dealers had no bargaining power, intermediation fees would equal 0 in every trade and dealers would be unable to cover their operation costs.) Our proof of existence of a nontrivial equilibrium for $\eta > 0$ relies on the properties of $\alpha(\cdot)$. As the measure of dealers becomes large, the instantaneous probability for a dealer to meet an investor is driven to zero, and given that the cost to participate in the market is strictly positive, the expected utility of a dealer becomes negative. Conversely, as the measure of dealers approaches 0, the rate at which a dealer meets an investor grows without bound and the expected profit of dealership becomes arbitrarily large. Consequently, since a dealer's expected profit is continuous in the contact rate, there is an intermediate value of $\upsilon$ such that expected profit equals 0.

of trades.

[30] Our free entry of dealers is analogous to the free entry of firms in Pissarides (2000). Rubinstein and Wolinsky (1987) also assume free entry of dealers (or *middlemen*), while Shevchenko (2004) has no entry but allows agents to choose whether to become middlemen.

Consider the limiting case where the dealer's operating cost, $\gamma$, tends to 0. Since the average fee $\bar{\phi}$ is positive and bounded away from 0 for any $\alpha < \infty$ (see the proof of Proposition 5), the free-entry condition (35) implies $\upsilon \to \infty$. This in turn implies that $\alpha \to \infty$, so the search equilibrium converges to the frictionless competitive equilibrium we characterized in Section 3. That is, $\{a_i\}$ satisfies (16) for $i = 1, ..., I$, $\phi_i(a) = 0$ for all $a$ and $i$, the equilibrium price, $p$, solves $\sum_i \pi_i u_i'^{-1}(rp) = A$, and the distribution of investors across asset holdings and preference types is $n_{ii} = \pi_i$ for each $i$, and $n_{ij} = 0$ for $j \neq i$.

Next, we analyze two cases where the equilibrium with entry is unique. First, suppose $\eta = 1$—an economy with monopolist dealers. From (34), the average fee is

$$\bar{\phi} = \sum_{i,j} \frac{\delta \pi_i \pi_j}{\alpha(\upsilon) + \beta + \delta} \frac{U_i(a_i) - U_i(a_j)}{r + \beta}. \tag{36}$$

The second factor in (36) is the expected discounted utility gain that an investor of type $i$ who holds portfolio $a_j$ gets from trading $a_i - a_j$. The average fee is a weighted average of each of these gains, with weights given by the proportion of agents of type $i$ who hold portfolio $j$ in the stationary distribution (the first factor in (36)). Since $\eta = 1$ implies that $\{U_i(\cdot)\}$ and $\{a_i\}$ are independent of $\alpha$, in this case the average fee only depends on $\alpha(\upsilon)$ through the distribution of investors, i.e., the "weights" in (36). As the number of dealers increases, a larger measure of investors hold their desired portfolios, which reduces dealers' opportunities to intermediate trades. Thus, in this case, $\bar{\phi}$ is strictly decreasing in $\alpha$ (and $\upsilon$). Therefore, the left side of (35) is strictly decreasing in $\upsilon$, which implies uniqueness of the steady-state equilibrium with entry.

From (35) with $\eta = 1$, we also obtain the following comparative static results: $d\upsilon/d\gamma < 0$, $d\upsilon/d\beta \gtrless 0$, $d\upsilon/d\delta \gtrless 0$. Higher operation costs reduce expected profits, so fewer dealers choose to operate. An increase in the rate at which investors get direct access to the asset market, $\beta$, is ambiguous in general because of two opposing effects. On the one hand, the fact that investors can trade more often without the need for a dealer reduces dealers' market power, and therefore also the average fee. This tends to reduce entry. On the other hand, as we saw in Section 6, with higher $\beta$ investors are able to rebalance their portfolios more frequently, and as a result they choose more extreme asset positions. For dealers, more extreme asset positions mean that on average they will earn higher intermediation fees. This tends to stimulate entry. The effect of an increase in $\delta$ on the measure of dealers is ambiguous for similar reasons.

In Figure 8 we report the behavior of the model with free entry and $\eta = 1$. Except for the value of $\eta$, the rest of the parametrization is as in the baseline without free entry. In addition, in this formulation we set $\alpha(\upsilon) = \upsilon^{0.75}$ and $\gamma = 0.0001$. The top-left panel displays the dealer's expected profits net of the flow cost, $\gamma$, illustrating the determination of the mass of dealers for the economy with $\beta = 0$. In the top-right panel, the solid line traces out the equilibrium measure of dealers as we vary $\beta$ from 0 to 5. For relatively low values of $\beta$, an increase in $\beta$ raises the measure of dealers. Over this range, allowing investors to have direct access to the market increases dealers' profits and makes them more willing to operate in the market. As a result, the rate at which investors contact dealers rises, as can be seen in the dashed line. For larger values of $\beta$, both the equilibrium measure of dealers and the rate at which investors contact dealers are decreasing in $\beta$. As a corollary, the bottom-left panel shows that for relatively low values of $\beta$, increases in $\beta$ can generate sharp reductions in the average execution delay, $\frac{1}{\alpha(\upsilon)+\beta}$. But for larger values of $\beta$, further increases in $\beta$ crowd out dealers and can cause an *increase* in the average execution delay.

The bottom-right panel of Figure 8 shows what happens to investors' optimal portfolios (for each level of the idiosyncratic shock $\varepsilon_i$) as $\beta$ varies from 0 to 5. When $\beta = 0$, investors are able to rebalance their portfolios very infrequently relative to the frequency of their idiosyncratic shocks.[31] In addition, since $\eta = 1$ in these experiments, investors get no surplus whenever they reallocate their portfolio through a dealer, which together with $\beta = 0$, implies that $\kappa = 0$. Hence, the optimal portfolio profile is extremely flat: when they make their portfolio allocation, investors do not vary their portfolio choice much as a function of their current preference shock. As $\beta$ rises, each investor's optimal portfolio profile becomes steeper: he chooses to hold large quantities of the asset when his current preference shock is high, and small ones when this shock is low. This increase in portfolio dispersion means that investors trade large volumes on average, and since fees are increasing in the size of the trade, this is what stimulates the entry of dealers.

Next, we consider a second case in which the measure of dealers is uniquely determined. Consider the limit as investors' preferences become linear, $u_i(a) \to \varepsilon_i a$. As in Section 4, let $\varepsilon_1 < \varepsilon_2 < ... < \varepsilon_I$, and recall that from (9), in this case only investors with the highest marginal utility want to hold assets. In this limit, the average fee becomes

$$\bar{\phi} = \frac{\eta\delta\left(\varepsilon_I - \bar{\varepsilon}\right) A}{\left[\beta + \delta + \alpha\left(\upsilon\right)\right]\left[r + \beta + \delta + \alpha\left(\upsilon\right)\left(1 - \eta\right)\right]}.$$

[31] The condition $\beta = 0$ means that investors can never access the market directly, and from the top-left panel we see that the equilibrium measure of dealers when $\beta = 0$ is very small, meaning that investors will have to wait a long time to be able to rebalance their portfolio through a dealer.

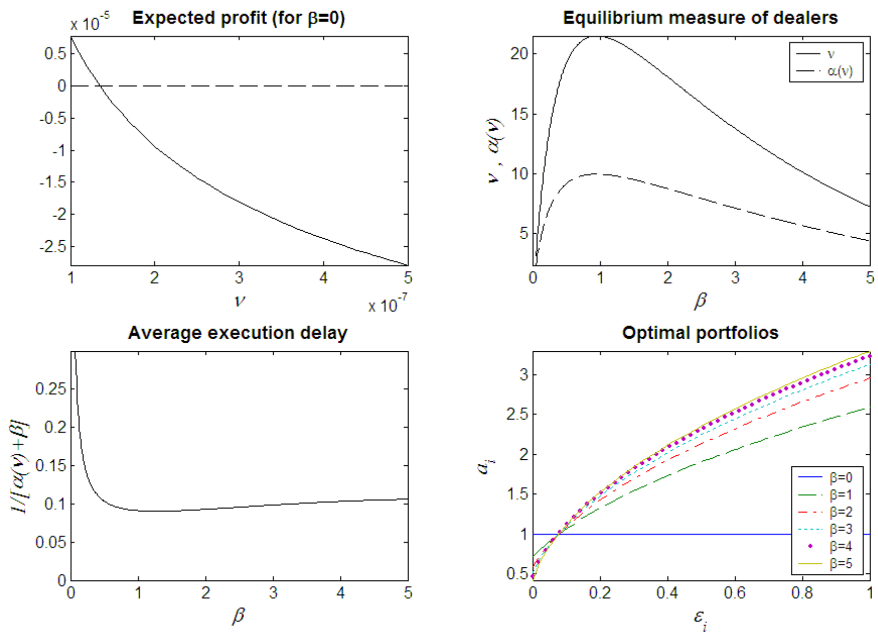


Figure 8: Model with entry of dealers and $\eta = 1$

The average fee is decreasing in $\upsilon$, so the left side of (35) is decreasing in $\upsilon$ and the equilibrium is unique. Furthermore, it is straightforward to verify that $\partial\upsilon/\partial\gamma < 0$, $\partial\upsilon/\partial\eta > 0$, $\partial\upsilon/\partial A > 0$, $\partial\upsilon/\partial r < 0$, $\partial\upsilon/\partial\beta < 0$ and $d\upsilon/d\delta \gtrless 0$. We discuss these results in turn.

Lower operation costs naturally imply more entry of dealers. Higher bargaining power for dealers means that they can extract a larger share from the gains from trade in a meeting with an investor, so the measure of dealers increases. Similarly, if the stock of assets increases, the size of each trade is larger and dealers make more profit. An increase in the discount rate reduces the number of dealers because investors benefit less from readjusting their portfolios, and therefore intermediation fees are lower. (Recall that with linear preferences, the gains from holding a portfolio are "backloaded," i.e., they materialize only when the portfolio is sold, and higher $r$ means these future gains are discounted more heavily.) An increase in $\beta$ discourages dealer entry because it strengthens the investor's outside option and also reduces the degree of mismatch between investors' desired and actual portfolios. Finally, an increase in the frequency of preference shocks has an ambiguous effect on the equilibrium measure of dealers. On the one hand, a higher $\delta$ generates more mismatch, which raises the return to intermediation. But on the other hand, since with larger $\delta$ preferences revert back to the mean marginal valuation $\bar{\varepsilon}$ faster, an increase in $\delta$ lowers the expected utility of the highest-valuation investor relative to the lower-valuation investors, which implies smaller gains from trade and consequently lower intermediation fees.

So far, we have discussed two special cases for which the equilibrium with entry is unique. But in general, the steady-state equilibrium with free entry need not be unique. An increase in the number of dealers leads to an increase in $\alpha(\upsilon)$. Faster trade means more competition among dealers, which tends to reduce intermediation fees. But as we have pointed out, an increase in $\alpha(\upsilon)$ also induces investors to take on more extreme asset positions (i.e., more in line with their current as opposed to the mean preference shock). This means that dealers will on average intermediate larger portfolio reallocations, which implies larger fees, since fees are increasing in the volume traded. The model will exhibit multiple steady states if the second effect is strong enough. (But for a given value $\upsilon$, the rest of the equilibrium, $\{(n_{ij}), (\phi_{ji}), (a_i), p\}$, is uniquely determined as in the previous sections.)[32]

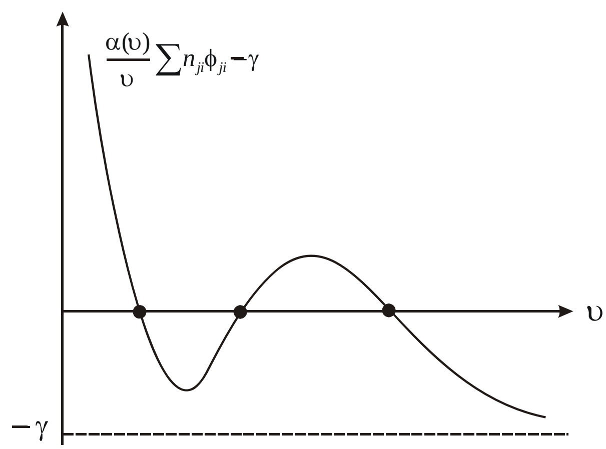


Figure 9: Multiple steady-state equilibria

In Figure 9 we provide a typical representation of a dealer's expected profit net of operation

[32] Notice that this type of multiplicity is new, and in fact, unlike the multiplicities in Diamond (1982) or Vayanos and Weill (2005), it is present even though we did not embed thick-market effects in our matching technology. That is, by assuming that $\partial[\alpha(\upsilon)/\upsilon] < 0$, we in fact assume that dealers reduce the rate at which other dealers contact investors. Thus, without the general equilibrium effect that operates through the shifts in portfolio compositions, our equilibrium would be unique. If instead we assumed that for some reason dealers find it easier to contact investors when more dealers participate in the market, i.e., $\partial[\alpha(\upsilon)/\upsilon] > 0$, then the model could display multiple equilibria even without the portfolio-composition effects that we have identified.

costs, $\frac{\alpha(\upsilon)}{\upsilon}\bar{\phi}-\gamma$, as a function of the measure of dealers. As $\upsilon$ approaches 0, the contact rate of dealers goes to infinity while $\bar{\phi}$ stays bounded away from 0. Therefore, dealers' expected profits are strictly positive for small $\upsilon$. As $\upsilon$ goes to infinity, the dealers' expected profits approach $-\gamma$. Thus, there are generically an odd number of steady-state equilibria. We typically find either one or three equilibria in our numerical work.[33] In case of multiple equilibria, the market can be stuck in a low-liquidity equilibrium where few dealers enter and investors engage in relatively small transactions. The low-liquidity equilibrium exhibits large spreads, small trade volume and long trade-execution delays.

The high and low equilibria share the following comparative statics: a decrease in the participation cost of dealers raises the measure of dealers in the market. If the decrease in the participation cost is large enough, the multiplicity of equilibria can be removed. (The expected profits curve in Figure 9 shifts upward.) Similarly, an increase in $\beta$ can eliminate the multiplicity of equilibria. Thus, it is possible that if the economy was initially in the low equilibrium, an increase in $\beta$ generates a large upward jump in the measure of dealers. So allowing investors to access the market directly need not crowd dealers out; in fact, it may even make intermediation more profitable.

To conclude the section, we want to argue that the multiplicity of steady-state equilibria is a robust feature of our model provided that the elasticity of the matching function $\alpha(\upsilon)$ is sufficiently close to one. This condition means that the marginal contribution of dealers to the matching process is large, and congestion effects on the dealer side are small. To see this clearly, consider a linear matching function, $\alpha(\upsilon)=\alpha_0\upsilon$, with $\alpha_0>0$. (This matching technology does not satisfy the assumption of strict concavity imposed earlier.) The rate at which dealers find orders to execute, $\alpha(\upsilon)/\upsilon$, is independent of the measure of dealers in the market (there are no congestion effects). From the free-entry condition, $\upsilon=0$ if $\alpha_0\bar{\phi}<\gamma$, $\upsilon=\infty$ if $\alpha_0\bar{\phi}>\gamma$ and $\upsilon\in[0,\infty]$ if $\alpha_0\bar{\phi}=\gamma$. If the average fee $\bar{\phi}(\upsilon)$ is hump-shaped, as it was in the top-right panel of Figure 6, for instance, the number of equilibria is either one or three. If there are multiple equilibria, then one of those equilibria is $\upsilon=0$, as illustrated in Figure 10. Furthermore, for the economy to reach the equilibrium with the highest number of dealers, there must be a critical mass of dealers: dealers enter only if the measure of dealers is above a threshold. By reducing the cost of dealership $\gamma$, or by improving the efficiency of the matching technology $\alpha_0$, one

[33] Note that the low and high steady states in Figure 9 are "stable" in the following heuristic sense: if one perturbates slightly the measure of dealers from its steady-state value, free entry tends to bring the measure of dealers toward its steady-state value.

can eliminate the multiplicity and reach a situation in which the equilibrium with the highest measure of dealers is the unique equilibrium.

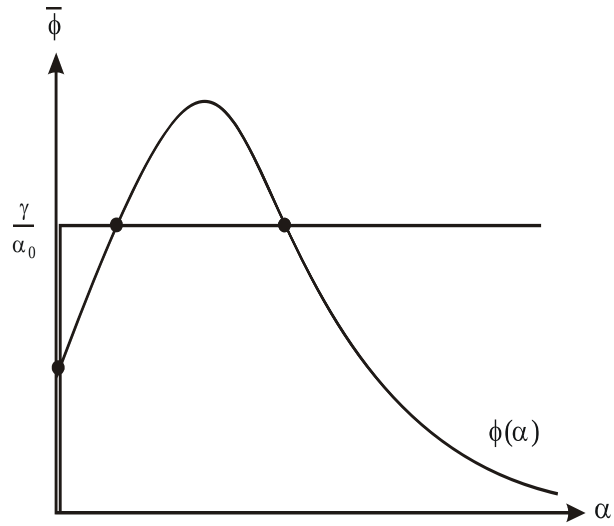


Figure 10: Free entry with linear matching technology

# 8 Efficiency

In this section we study the problem of a social planner who maximizes the expected discounted sum of all agents' utilities. When choosing allocations, the planner is subject to the same frictions that investors and dealers face in the decentralized formulation studied in the previous sections. Specifically, these frictions imply that over a small interval of time of length $dt$, the planner can only reallocate assets among a measure $(\alpha + \beta)\, dt$ of investors chosen at random from the population. We will study efficiency in both the model with a fixed number of dealers and the model with free entry of dealers.

## 8.1 Investors' portfolios

Let $H_t(a, i)$ denote the distribution of investors across portfolios and preference types at time $t$. Since at any point in time all investors access the market according to independent stochastic processes with identical distributions, the measure of assets that can be reallocated among the $\alpha+\beta$ randomly drawn investors is $(\alpha + \beta) \int a dH_t(a, i) = (\alpha + \beta)\, A$. Thus, the quantity of assets that can be reallocated among investors depends only on the mean of $H_t(a,i)$. Consequently, the planner's decision of how to allocate assets at time $t$ affects neither the measure of investors he will draw in the future nor the total measure of assets that these investors hold. In other words, $H_t(a,i)$ is not a state variable for the planner's problem.

The planner chooses among allocations $\{a_i(t)\}_{i=1}^I$ that specify how to distribute the measure $(\alpha+\beta)\,A$ of assets among the measure $\alpha+\beta$ of investors whose portfolios he can reallocate at date $t$. Let $\tilde{V}_i(a)$ denote the expected discounted utility of an investor of type $i$ who holds stock of assets $a$ from the present until the next time he gains access to the asset market, i.e.,

$$\tilde{V}_i(a) = \mathbb{E}_t\left[\int_t^{t+T} e^{-r(\tau-t)} u_{k(\tau)}(a)d\tau\right]. \tag{37}$$

The expectation is over the random variables $T$ and $k$, where $T$ is the time until the investor regains access to the asset market and $k(\tau)$ is the preference type at time $\tau$, which evolves according to the stochastic process for idiosyncratic preference shocks. Note that—since the stochastic processes for both random variables are stationary—the right-hand side of (37) is independent of calendar time, $t$. The value function $\tilde{V}_i(a)$ satisfies the following flow Bellman equation:

$$r\tilde{V}_i(a) = u_i(a) + \delta\sum_j \pi_j\left[\tilde{V}_j(a) - \tilde{V}_i(a)\right] - (\alpha+\beta)\tilde{V}_i(a), \tag{38}$$

and in turn, (38) implies

$$\tilde{V}_i(a) = \frac{(r+\alpha+\beta)\,u_i(a) + \delta\sum_j \pi_j u_j(a)}{(r+\alpha+\beta+\delta)\,(r+\alpha+\beta)}. \tag{39}$$

Since general goods enter linearly in the utility function of all agents, the consumption and production of those goods net out to 0 and can be ignored by the planner. Therefore, the planner only maximizes the investors' direct utilities from holding the asset. Given an initial distribution $H_0\,(a,i)$ of investors over asset holdings and preference types, the planner solves

$$\max_{\{a_i(t)\}}\left\{\Psi_0 + \int_0^\infty \sum_i e^{-rt}\,(\alpha+\beta)\,n_i(t)\tilde{V}_i\,[a_i(t)]\,dt\right\}$$

$$\text{s.t. } \dot{n}_i(t) = \delta\,[\pi_i - n_i(t)]\,,\ \text{for } i = 1,...,I \tag{40}$$

$$\sum_i (\alpha+\beta)\,n_i(t)a_i(t) \le (\alpha+\beta)\,A, \tag{41}$$

where $\Psi_0 \equiv \int_i \tilde{V}_i(a)dH_0(a,i)$. The term $\Psi_0$ captures the utility of all investors before they access the marketplace for the first time. This term is a constant because the planner can only reallocate assets among the $\alpha+\beta$ randomly drawn investors who contact the marketplace. The second term in the objective function states that over an interval of time of length $dt$, there is a measure $(\alpha+\beta)\,n_i(t)dt$ of investors of type $i$ who can have their portfolios rebalanced. An investor of type $i$ is assigned a portfolio $a_i(t)$. The planner's choices are constrained by the law of motion of the measure of investors of each preference type, (40), and must satisfy the resource constraint (41). The following proposition characterizes the optimal allocation and summarizes the efficiency properties of the equilibrium of the model with a fixed measure of dealers.

**Proposition 6** *Consider the economy with an exogenous measure of dealers. Then:*

*(a) The efficient allocation* $\{a_i(t)\}_{i=1}^{I}$ *satisfies*

$$\frac{(r+\alpha+\beta)\,u_i'\left[a_i(t)\right]+\delta\sum_k \pi_k u_k'\left[a_i(t)\right]}{r+\alpha+\beta+\delta}\leq\lambda(t),\qquad \text{“}=\text{”}\quad \textit{if } a_i(t)>0, \tag{42}$$

*where* $\lambda(t)$ *is the Lagrange multiplier on the resource constraint (41) scaled up by* $(r+\alpha+\beta)$*. Let* $a_i^*\left[\lambda(t)\right]$ *denote the solution to (42), then the shadow price of an asset,* $\lambda(t)$*, satisfies*

$$\sum_i\left[(1-e^{-\delta t})\pi_i+e^{-\delta t}n_i(0)\right]a_i^*\left[\lambda(t)\right]=A \tag{43}$$

*at each date* $t$*.*

*(b) The equilibrium is efficient if and only if* $\eta=0$*.*

The equilibrium with bargaining is efficient if and only if dealers have no bargaining power. This result may seem surprising given that the Nash solution implies efficient trade for each investor-dealer match. Specifically, recall that in the equilibrium, the portfolio implied by the bargaining outcome maximizes the joint surplus of an investor and a dealer, and it is in fact the same portfolio that an investor would choose if he had *direct* access to the market. The inefficiency arises from a standard holdup problem due to ex-post bargaining. When conducting a trade, investors anticipate the fact that they will have to pay fees for rebalancing their portfolios in the future and that these intermediation fees increase with the surplus that those future trades generate. As a result, at the margin, investors are discouraged from taking positions that tend to lead to large portfolio reallocations in the future. This inefficiency of the equilibrium with bargaining gets mitigated as $\alpha$ or $\beta$ increases, since subjecting dealers to increased competition reduces transaction fees.

## 8.2 Entry of dealers

Finally, we investigate the efficiency properties of equilibrium with free entry of dealers. The dynamic planner's problem is now much more complex since the measure of dealers at any point in time will typically depend on the whole distribution $H_t(a,i)$, not just its mean. To keep the analysis manageable, here we consider the case where the discount rate is close to 0, i.e., we characterize the allocation chosen by a social planner who maximizes steady-state welfare. In this case, the planner solves

$$\begin{aligned} \max_{\{a_i\},\upsilon} \sum_{i,j} n_{ij} u_j(a_i) - \upsilon\gamma \\ \text{s.t.} \sum_{i,j} n_{ij} a_i = A, \end{aligned} \tag{44}$$

where the steady-state distribution $\{n_{ij}\}$ satisfies (13) and (14). The planner maximizes the population-weighted sum of investors' utilities from holding the asset, net of the participation costs of the dealers and taking into account that the stationary distribution $\{n_{ij}\}$ depends on the measure of dealers, $\upsilon$.

**Proposition 7** *Consider the economy with free entry of dealers and let $r \approx 0$. Then:*

*(a) The efficient allocation $\{(a_i)_{i=1}^I, \upsilon\}$ satisfies*

$$\frac{(\alpha+\beta)\, u_i'(a_i) + \delta \sum_k \pi_k u_k'(a_i)}{\alpha+\beta+\delta} \le \lambda, \qquad \text{“}=\text{”} \quad \text{if } a_i > 0, \tag{45}$$

$$\frac{\alpha'(\upsilon)}{\alpha(\upsilon)+\beta+\delta} \sum_{i,j} \frac{\delta \pi_i \pi_j \left[u_i(a_i) - u_j(a_i)\right]}{\alpha(\upsilon)+\beta+\delta} = \gamma, \tag{46}$$

*and (44), where $\lambda$ is the Lagrange multiplier on the resource constraint (44) scaled up by $(\alpha+\beta)$.*

*(b) An equilibrium with free entry is efficient if and only if $\eta = 0$ and $\alpha'(\upsilon)\upsilon/\alpha(\upsilon) = \eta$.*

As before, investors' portfolios are efficient if and only if dealers have no bargaining power. Entry introduces an additional inefficiency: when a dealer enters the market, he imposes a negative externality on other dealers' order flow. As it is well-known since Hosios (1990), these search externalities are internalized if and only if the elasticity of the matching function coincides with dealers' bargaining power. If the Hosios condition $\alpha'(\upsilon)\upsilon/\alpha(\upsilon) = \eta$ holds, then the equilibrium allocations can be made arbitrarily close to the efficient allocations by making $\eta$ arbitrarily close to 0. But there is no free-entry equilibrium with $\eta = 0$, so an equilibrium with entry is always inefficient.

# 9 Conclusion

We have developed a simple search-theoretic model of the exchange process in an asset market. The asset market we modeled captures the salient features of many financial trades in various contexts—and in particular of those carried out in over-the-counter markets. A fraction of trades is intermediated by dealers who have access to an interdealer market, while the rest of the trades are conducted directly between investors. In both cases finding a counterpart to execute the trade entails delays. We have examined how these trading delays affect the level and volatility of asset prices, the size of bid-ask spreads, the volume of trade, the allocation of assets across investors and the profitability and participation decisions of dealers. As far as we know, our analysis is one of the first theoretical attempts to study the positive and normative implications of the introduction of technological innovations in trading—such as electronic and automated trading—that have increased the speed at which financial transactions are matched and executed.

From a methodological point of view, we have generalized the model of Duffie, Gârleanu and Pedersen (2005) along several dimensions. We have relaxed the portfolio restrictions so that investors can hold any nonnegative quantity of asset, extended the model to consider more general preferences and more general forms of investor heterogeneity, allowed for idiosyncratic as well as aggregate uncertainty, and granted investors direct, as well as indirect (i.e., dealer-intermediated), access to a competitive market. We have also endogenized the provision of liquidity by endogenizing the measure of dealers.

In terms of findings, we have shown that the level and volatility of the asset price need not be affected by the degree of the trading frictions; and that if they are, the sign of the effect depends on the curvature of the utility function. We have found that a reduction in trading delays can increase the dealers' average profit despite the fact that, for a given trade size, intermediation fees decrease with a reduction in trading delays. We have also shown that there can be multiple equilibria in a version of the model where the measure of dealers is endogenous. Equilibria with high asset turnover, narrow bid-ask spreads and high level of participation of dealers can coexist with equilibria with low turnover, narrow spreads and longer trading delays. We have also studied the model from a normative standpoint and found that generically, equilibria are inefficient. Investors' portfolio choices are inefficient because of a holdup problem in their relationships with dealers. Entry of dealers is inefficient because of a standard search externality.

By way of numerical simulations, we have shown that while the model is stylized, it allows for fairly general forms of investor heterogeneity and it has relatively few parameters that map naturally into observables. One could imagine calibrating or estimating the model using data on trade execution in over-the-counter markets. We think that much could be learned from such exercises. For example, one could quantify the welfare gains associated with a given reduction in trading frictions, or predict the impact that the introduction of electronic trading networks will have on bid-ask spreads, average execution times, trade volume and other standard measures of market liquidity. Various extensions are worth considering. First, there are many issues, such as the dynamic provision of liquidity by dealers who can hold asset positions, that would require a more detailed study of the model dynamics. Second, as an alternative to bilateral bargaining, one could explore alternative trading mechanisms that combine price-posting and directed search. Finally, a model with multiple assets could be used to study how the various assets' liquidity properties are jointly determined.

## References


[1] Allen, Helen, John Hawkins and Setsuya Sato (2001). "Electronic trading and its implications for financial systems," *BIS Papers* 7, 30–52.

[2] Barclay, Michael J., William G. Christie, Jeffrey H. Harris, Eugene Kandel and Paul H. Schultz (1999). "Effects of market reform on the trading costs and depths of NASDAQ stocks," *Journal of Finance* 65(1) (February): 1–34.

[3] Boehmer, Ekkehart (2005). "Dimensions of execution quality: Recent evidence for U.S. equity markets," *Journal of Financial Economics* 78, 553–582.

[4] Boehmer, Ekkehart, Robert Jennings and Li Wei (2005). "Public disclosure and private decisions: Equity market execution quality and order routing," Unpublished working paper.

[5] Diamond, Peter A. (1982). "Aggregate demand management in search equilibrium," *Journal of Political Economy* 90(5) (October): 881–894.

[6] Duffie, Darrell, Nicolae Gârleanu and Lasse Heje Pedersen (2005). "Over-the-counter markets," *Econometrica* 73(6) (November): 1815–1847.

[7] Edwards, Amy K., Lawrence Harris and Michael Piwowar (2004). "Corporate bond market transparency and transaction costs," mimeo, Social Science Research Network.

[8] Gale, Douglas (1987). "Limit theorems for markets with sequential bargaining," *Journal of Economic Theory* 43, 20–54.

[9] Gârleanu, Nicolae (2005). "Portfolio choice and pricing in illiquid markets," mimeo, University of Pennsylvania Wharton School.

[10] Harris, Larry (2003). *Trading and exchanges: Market microstructure for practitioners.* New York: Oxford University Press.

[11] He, Hua, and David M. Modest (1995). "Market frictions and consumption-based asset pricing," *Journal of Political Economy* 103, 94–117.

[12] Heaton, John, and Deborah J. Lucas (1995). "The importance of investor heterogeneity and financial market imperfections for the behavior of asset prices," *Carnegie–Rochester Conference Series on Public Policy* 42, 1–32.

[13] Hosios, Arthur J. (1990). "On the efficiency of matching and related models of search and unemployment," *Review of Economic Studies* 57, 279–298.

[14] Joys, Jonathan (2001). "Electronic trading systems and fixed income markets: 2001," Tuck School of Business at Dartmouth, Glassmeyer/McNamee Center for Digital Strategies, CDS Case #6-0006.

[15] Kiyotaki, Nobuhiro, and Randall Wright (1993). "A search-theoretic approach to monetary economics," *American Economic Review* 83, 63–77.

[16] Lo, Andrew W., Harry Mamaysky and Jiang Wang (2004). "Asset prices and trading volume under fixed transactions costs," *Journal of Political Economy* 112, 1054–1090.

[17] Luttmer, Erzo G. J. (1996). "Asset pricing in economies with frictions," *Econometrica* 64, 1439–1467.

[18] McAndrews, James and Chris Stefanadis (2000). "The emergence of electronic communication networks in the U.S. equity markets," *Current Issues in Economics and Finance of the Federal Reserve Bank of New York* 6, 1–6.

[19] Miao, Jianjun (2006). "A search model of centralized and decentralized trade," *Review of Economic Dynamics* 9(1) (January): 68–92.

[20] NASDAQ (2000). "Market quality statistics." http://www.marketdata.nasdaq.com/mr6d.html (March).

[21] Pissarides, Christopher A. (2000). *Equilibrium unemployment theory*, 2nd ed. Cambridge, MA: MIT Press.

[22] Rubinstein, Ariel and Asher Wolinsky (1987). "Middlemen," *Quarterly Journal of Economics* 102, 581–594.

[23] Rust, John and George Hall (2003). "Middlemen versus market makers: A theory of competitive exchange," *Journal of Political Economy* 111, 353–403.

[24] Saunders, Anthony, Anand Srinivasan and Ingo Walter (2002). "Price formation in the OTC corporate bond markets: A field study of the interdealer market," *Journal of Economics and Business* 54, 95–113.

[25] Schultz, Paul (2001). "Corporate bond trading costs: A peek behind the curtain," *Journal of Finance* 56, 677–698.

[26] Shevchenko, Andrei (2004). "Middlemen," *International Economic Review* 45, 1–24.

[27] Spulber, Daniel (1996). "Market making by price-setting firms," *Review of Economic Studies* 63, 559–580.

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

[29] Vayanos, Dimitri (1998). "Transaction costs and asset prices: A dynamic equilibrium model," *Review of Financial Studies* 11, 1–58.

[30] Vayanos, Dimitri and Tan Wang (2002). "Search and endogenous concentration of liquidity in asset markets," unpublished manuscript.

[31] Vayanos, Dimitri and Pierre-Olivier Weill (2005). "A search-based theory of the on-the-run phenomenon," mimeo, London School of Economics and NYU Stern School of Business.

[32] Wahal, Sunil (1997). "Entry, exit, market makers, and the bid-ask spread," *Review of Financial Studies* 10(3) (Fall): 871–901.

[33] Weill, Pierre-Olivier (2005a). "Leaning against the wind," unpublished manuscript.

[34] Weill, Pierre-Olivier (2005b). "Liquidity premia in dynamic bargaining markets," unpublished manuscript.

[35] Weston, James (2000). "Competition on the NASDAQ and the impact of recent market reforms," *Journal of Finance* 55(6) (December): 2565–2598.

[36] Weston, James (2002). "Electronic communication networks and liquidity on the NASDAQ," *Journal of Financial Services Research* 22, 125–139.

# A Appendix

**Proof of Lemma 1.** The Nash solution requires the outcome to be Pareto efficient. Since agents' payoffs are linear in $\phi$, $a^b$ must maximize the surplus from the match, namely $V_i\left(a^b\right) - V_i(a) - p\left(a^b - a\right)$. Differentiating the Nash product in (3) with respect to $\phi$ and equating to zero gives (5). ■

**Proof of Lemma 2.** Let $S = \mathbb{R}_+ \times \{1, ..., I\}$, $\mathcal{C} = \{g : S \to \mathbb{R} \mid g(a,i)$ is continuous in $a$ and bounded above$\}$ and $\mathcal{C}' = \{f : S \to \mathbb{R} \mid f(a,i) = g(a,i) + pa$, for some $g \in \mathcal{C}\}$. Define $v(a,i) \equiv \frac{U_i(a) - rpa}{r+\beta+\alpha(1-\eta)} + pa$. Note that our assumptions on $u_i$ imply that $v(a,i) \in \mathcal{C}'$. Next, rewrite (6) as

$$V_i(a) = \frac{U_i(a) + [\beta + \alpha(1-\eta)](pa + \bar{V}_i)}{r + \beta + \alpha(1-\eta)}, \tag{47}$$

where

$$\bar{V}_i = \frac{r + \beta + \alpha(1-\eta)}{r + \beta + \delta + \alpha(1-\eta)} \max_x [V_i(x) - px] + \frac{\delta}{r + \beta + \delta + \alpha(1-\eta)} \textstyle\sum_k \pi_k \max_x [V_k(x) - px]$$

for $i = 1, ..., I$. The right-hand side of (47) defines an operator $T$:

$$T(V)(a,i) = \max_{(x_k)_{k=1}^I} \{v(a,i) + \hat{\beta}[(1-\hat{\delta})[V(x_i,i) - px_i] + \hat{\delta}\textstyle\sum_k \pi_k [V(x_k,k) - px_k]]\}, \tag{48}$$

with $\hat{\beta} = \frac{\beta+\alpha(1-\eta)}{r+\beta+\alpha(1-\eta)}$, and $\hat{\delta} = \frac{\delta}{r+\beta+\delta+\alpha(1-\eta)}$. We wish to show there exists a unique solution $V(a,i)$ to $(TV)(a,i) = V(a,i)$, and that $V(a,i) = V_i(a)$, with $V_i(a)$ as in (8). Suppose $V \in \mathcal{C}'$, then the maximization on the right-hand side of (48) has a solution, $(a_j)_{j=1}^I$, and this solution is independent of $a$. Thus, $T(V)(a,i) = v(a,i) + \hat{\beta}c_i$, where $c_i = (1-\hat{\delta})[V(a_i,i) - pa_i] + \hat{\delta}\sum_k \pi_k [V(a_k,k) - pa_k]$ is a constant. Therefore $T : \mathcal{C}' \to \mathcal{C}'$. Consider the metric space $(\mathcal{C}', \|\cdot\|)$, where $\|\cdot\|$ denotes the *sup norm*. We next show that $T$ is a contraction on $(\mathcal{C}', \|\cdot\|)$. Consider an arbitrary pair $V^1, V^2 \in \mathcal{C}'$ and let $(a_k^j)_{k=1}^I \in \arg\max_{(x_k)_{k=1}^I} \{(1-\hat{\delta})[V^j(x_i,i) - px_i] + \hat{\delta}\sum_k \pi_k \left[V^j(x_k,k) - px_k\right]\}$ for $j = 1, 2$. Fix $(a,i) \in S$, then

$$\begin{aligned}
TV^1(a,i) - TV^2(a,i) &= \hat{\beta}(1-\hat{\delta})\left[V^1\left(a_i^1,i\right) - V^2\left(a_i^2,i\right) - \left(pa_i^1 - pa_i^2\right)\right] \\
&\quad + \hat{\beta}\hat{\delta}\sum_k \pi_k \left[V^1\left(a_k^1,k\right) - V^2\left(a_k^2,k\right) - \left(pa_k^1 - pa_k^2\right)\right] \\
&\le \hat{\beta}(1-\hat{\delta})\left[V^1\left(a_i^1,i\right) - V^2\left(a_i^1,i\right)\right] + \hat{\beta}\hat{\delta}\textstyle\sum_k \pi_k \left[V^1\left(a_k^1,k\right) - V^2\left(a_k^1,k\right)\right] \\
&\le \hat{\beta} \sup_{i \in \{1,...,I\}} \left[V^1\left(a_i^1,i\right) - V^2\left(a_i^1,i\right)\right] \le \hat{\beta} \sup_{(a,i) \in S} \left[V^1(a,i) - V^2(a,i)\right] \\
&\le \hat{\beta} \sup_{(a,i) \in S} \left|V^1(a,i) - V^2(a,i)\right| = \hat{\beta}\left\|V^1 - V^2\right\|.
\end{aligned}$$

Similarly, we can obtain $TV^2(a,i) - TV^1(a,i) \leq \hat{\beta} \left\| V^1 - V^2 \right\|$. Hence,

$$\left| TV^1(a,i) - TV^2(a,i) \right| \leq \hat{\beta} \left\| V^1 - V^2 \right\|, \quad \forall (a,i) \in S.$$

Taking the *sup* over $(a,i)$ on the left-hand side of this inequality, we get $\left\| TV^1 - TV^2 \right\| \leq \hat{\beta} \left\| V^1 - V^2 \right\|$. Since $\hat{\beta} \in [0,1)$, $T$ is a contraction with modulus $\hat{\beta}$ on $\mathcal{C}'$. Since $(\mathcal{C}', \|\cdot\|)$ is complete (because $(\mathcal{C}, \|\cdot\|)$ is complete), it follows from the Contraction Mapping Theorem (e.g., Stokey and Lucas (1989), Theorem 3.2) that $T$ has a unique fixed point $V \in \mathcal{C}'$. It is a matter of algebra to use (48) to verify that indeed $(TV)(a,i) = V(a,i)$ for $V(a,i) = V_i(a)$ given by (8). ■

**Proof of Lemma 3.** From (11) and (12),

$$\delta \pi_j n_{i\cdot} - (\alpha + \beta + \delta) n_{ij} = 0, \quad \text{for } j \neq i \tag{49}$$

$$\delta \pi_i n_{i\cdot} + (\alpha + \beta) n_{\cdot i} - (\alpha + \beta + \delta) n_{ii} = 0, \tag{50}$$

where $n_{i\cdot} = \sum_k n_{ik}$ and $n_{\cdot i} = \sum_k n_{ki}$ are the marginal distributions. Sum (49) over $j$, and add (50) to the resulting expression to get $n_{i\cdot} = n_{\cdot i}$. Then sum (49) over $i$, and add the resulting expression to $\delta \pi_j n_{j\cdot} + (\alpha + \beta) n_{\cdot j} = (\alpha + \beta + \delta) n_{jj}$ (this is (50), but with $i = j$) to get $n_{\cdot j} = \pi_j$. Thus, $n_{\cdot j} = n_{j\cdot} = \pi_j$. Substituting these marginals into (49) and (50) yields (13) and (14). ■

**Proof of Proposition 1.** The steady-state distribution $(n_{ij})_{i,j=1}^{I}$ is unique and given by (13) and (14). From (9), any interior portfolio choice $a_i$ is a strictly decreasing function of $p$ for every $i$. Therefore, the market-clearing condition (15) determines a unique $p$. Given $p$, there is a unique $a_i$ that solves (9). Finally, given $p$ and $a_i$, (10) gives the fee $\phi_i(\cdot)$ for each $i$. ■

**Proof of Lemma 4.** From (17), we see that the model is formally equivalent to one where investors have *effective* direct access to the asset market with Poisson rate $\kappa$. Using this observation, the $V_i^s(a)$ from (17) also satisfies

$$V_i^s(a) = \mathbb{E} \left\{ \int_0^T e^{-rt} u_{k(t)}^{w(t)}(a) dt + e^{-rT} \max_{a'} [V_{k(T)}^{w(T)}(a') - p^{w(T)}(a' - a)] \right\}, \tag{51}$$

where $w(t) \in \{H, L\}$ denotes the aggregate state, and $k(t) \in \{1, 2, ..., I\}$ the investor's preference type at time $t$. In (51), $T$ is an exponentially distributed random variable with mean $1/\kappa$, that denotes the period of time that elapses until the investor gains direct effective access to the market. The expectations operator, $\mathbb{E}$, is with respect to the random variable $T$ and the two independent Poisson processes $\{w(x), k(x)\}$, and is conditional on $(w(0), k(0)) = (s, i)$. We now proceed to simplify (51). First, notice that (51) can be rewritten as

$$V_i^s(a) = \tilde{V}_i^s(a) + \mathbb{E}[e^{-rT}p^{w(T)}]a + \Delta_i^s, \tag{52}$$

where $\tilde{V}_i^s(a) = \mathbb{E}\int_0^T e^{-rx}u_{k(x)}^{w(x)}(a)dx$, and $\Delta_i^s = \mathbb{E}e^{-rT}\max_{a'}[V_{k(T)}^{w(T)}(a') - p^{w(T)}a']$. Using (52), the optimization problem of an investor, $\max_{a_i^s}[V_i^s(a_i^s) - p^s a_i^s]$, can be written as

$$\max_{a'}\{\tilde{V}_i^s(a') - [p^s - \mathbb{E}(e^{-rT}p^{w(T)})]a'\}. \tag{53}$$

The expectation in (53) is over the random variables $T$ and $w(T)$, conditional on the current aggregate state, $s$. Next, we proceed in two steps: $(i)$ derive a simpler expression for $\tilde{V}_i^s(a)$, and $(ii)$ show how to simplify $\mathbb{E}_s\left[e^{-rT}p^{w(T)}\right]$.

$(i)$. The value $\tilde{V}_i^s(a)$ satisfies the following Bellman equation

$$r\tilde{V}_i^s(a) = u_i^s(a) + \delta\sum_k \pi_k[\tilde{V}_k^s(a) - \tilde{V}_i^s(a)] + \lambda^s[\tilde{V}_i^{s'}(a) - \tilde{V}_i^s(a)] - \kappa\tilde{V}_i^s(a).$$

After some manipulations, we find

$$\tilde{V}_i^s(a) = \frac{U_i^s(a)}{r+\kappa}. \tag{54}$$

$(ii)$. The expected discounted resale price of the asset satisfies

$$\mathbb{E}_s[e^{-rT}p^{w(T)}] = \frac{\kappa}{r+\kappa}\left[\frac{(r+\kappa+\lambda^{s'})p^s + \lambda^s p^{s'}}{r+\kappa+\lambda}\right], \tag{55}$$

for $s \in \{H, L\}$, and $s' \in \{H, L\} \setminus \{s\}$. Substituting (54) and (55), (53) is equivalent to

$$\max_{a_i^s}\{U_i^s(a_i^s) - [rp^s - \frac{\kappa\lambda^s}{r+\kappa+\lambda}(p^{s'} - p^s)]a_i^s\},$$

which is identical to (18) once we let $rp^s - \frac{\kappa\lambda^s}{r+\kappa+\lambda}(p^{s'} - p^s) = \xi^s$. ■

**Proof of Proposition 2.** Differentiating (15) we obtain

$$\frac{dp}{d\alpha} = \frac{\sum_i \pi_i \partial a_i/\partial\alpha}{-\sum_i \pi_i \partial a_i/\partial p}.$$

From (25), we know that the denominator of this expression is strictly positive, so we focus on the sign of the numerator. Differentiating (25) to obtain $\partial a_i/\partial\alpha$, multiplying by $\pi_i$, and adding over all $i$, we arrive at

$$\sum_i \frac{\partial a_i}{\partial\alpha}\pi_i = \frac{(1-\eta)\delta}{(r+\delta+\kappa)^2 rp}\sum_i \frac{[u'(a_i)]^2}{-u''(a_i)}(\varepsilon_i - \bar{\varepsilon})\pi_i.$$

Suppose $-[u'(a)]^2/u''(a)$ is strictly increasing in $a$. Let $\bar{a}$ denote the $a$ that solves (25) for $\bar{\varepsilon}_i = \bar{\varepsilon}$. Then, note that $-[u'(a_i)]^2(\varepsilon_i - \bar{\varepsilon})/u''(a_i) \geq -[u'(\bar{a})]^2(\varepsilon_i - \bar{\varepsilon})/u''(\bar{a})$ for each $i$, with strict inequality for all $i$ such that $\varepsilon_i \neq \bar{\varepsilon}$. Thus, $\sum_i \frac{\partial a_i}{\partial \alpha}\pi_i > 0$ and consequently, $\frac{dp}{d\alpha} > 0$. Similar reasoning implies $\frac{dp}{d\alpha} < 0$ if $-[u'(a)]^2/u''(a)$ is strictly decreasing and $\frac{dp}{d\alpha} = 0$ if $-[u'(a)]^2/u''(a)$ is constant in $a$. ■

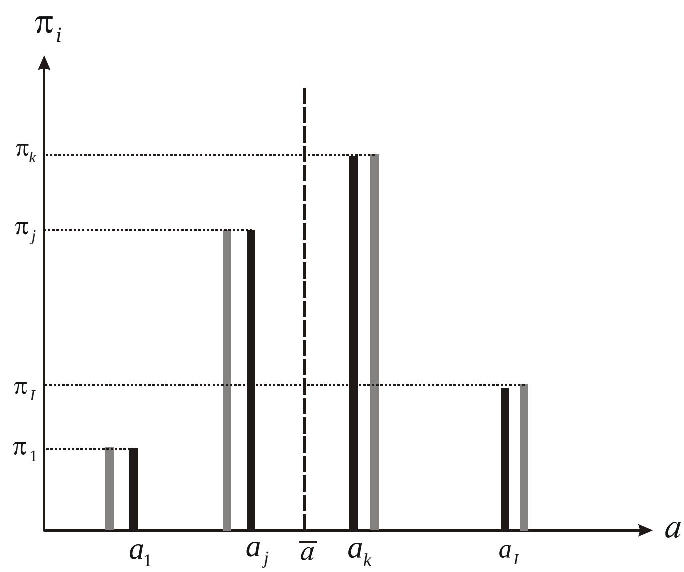


Figure 11: Distribution of asset holdings

**Proof of Proposition 3**. Let $a_i(\kappa)$ denote the individual demand for the asset by an agent whose current preference shock is $\varepsilon_i$ in an economy where the direct effective access rate to the asset market is $\kappa < \kappa'$. From (26), for each $i$,

$$a_i(\kappa) = \frac{A}{\sum_j \pi_j \left[\frac{(r+\kappa)\varepsilon_j + \delta\bar{\varepsilon}}{(r+\kappa)\varepsilon_i + \delta\bar{\varepsilon}}\right]^{1/\sigma}}.$$

Then, one can verify that there exists a unique $\tilde{\varepsilon} \in (\varepsilon_1, \varepsilon_I)$ such that $a_i(\kappa') > a_i(\kappa)$ for all $\varepsilon_i > \tilde{\varepsilon}$, $a_i(\kappa') < a_i(\kappa)$ for all $\varepsilon_i < \tilde{\varepsilon}$ and $a_i(\kappa') = a_i(\kappa) \equiv \tilde{a}$ if $\varepsilon_i = \tilde{\varepsilon}$. With Lemma 3, the cumulative distribution of assets across investors for the economy indexed by $\kappa$, is $\mathbb{G}_\kappa(a) = \sum_j \mathbf{1}_{\{a_j(\kappa) \leq a\}}\pi_j$. This, and the fact that $(\kappa' - \kappa)[a_i(\kappa') - a_i(\kappa)] > 0$ iff $\varepsilon_i > \tilde{\varepsilon}$ implies that $\mathbb{G}_{\kappa'}(a) \geq \mathbb{G}_\kappa(a)$ for all $a < \tilde{a}$ and $\mathbb{G}_{\kappa'}(a) \leq \mathbb{G}_\kappa(a)$ for all $a > \tilde{a}$. Thus, given that both densities have the same mean and $a_I(\kappa') > a_I(\kappa)$, the fact that the cumulative density functions cross only once implies that $\mathbb{G}_\kappa$ dominates $\mathbb{G}_{\kappa'}$ in the second-order stochastic sense. (Propostion 3 is illustrated in Figure 11, where we represent the distribution associated with $\kappa$ in black and the distribution associated with $\kappa'$ in grey.) ■

**Proof of Proposition 4**. Differentiating, we find

$$\frac{\partial}{\partial a}\left[\frac{\phi_i(a)}{a_i - a}\right] = \frac{\eta}{r+\beta+\alpha(1-\eta)}\left[\frac{U_i(a_i) - U_i(a) - U_i'(a)(a_i - a)}{(a_i - a)^2}\right] \leq 0.$$

Thus, the intermediation fee per unit of asset traded is increasing in the size of the trade if and only if $U_i$ is concave. ■

**Proof of Proposition 5.** Using (7), we can write (34) as

$$\bar{\phi} = \frac{\eta\delta}{(\beta+\delta+\alpha)[r+\beta+\delta+\alpha(1-\eta)]}\sum_{i,j}\pi_i\pi_j[u_i(a_i) - u_i(a_j)].$$

From (9) we know that $a_i$ is a continuous function of $\upsilon$ and $p$, i.e., $a_i = a_i(\upsilon, p)$. From (15), there is a unique $p$ that clears the asset market and it is a continuous function of $\upsilon$, i.e., $p = p(\upsilon)$. Thus, $a_i = a_i[\upsilon, p(\upsilon)]$ is a continuous function of $\upsilon$. Define the map $\Gamma(\upsilon)$ as

$$\Gamma(\upsilon) \equiv \frac{[\alpha(\upsilon)/\upsilon]\,\delta\eta\sum_{i,j}\pi_i\pi_j\{u_i[a_i(\upsilon)] - u_i[a_j(\upsilon)]\}}{[\beta+\delta+\alpha(\upsilon)][r+\beta+\delta+(1-\eta)\alpha(\upsilon)]}. \tag{56}$$

This is the left-hand side of the free-entry condition (35). First, we establish that $\lim_{\upsilon\to 0}\Gamma(\upsilon) = \infty$. Recall that

$$a_i = \arg\max_a\left[\frac{(r+\kappa)u_i(a)+\delta\sum_k\pi_k u_k(a)}{r+\delta+\kappa} - rpa\right],$$

therefore,

$$\frac{(r+\kappa)\,u_i(a_i)+\delta\sum_k\pi_k u_k(a_i)}{r+\delta+\kappa} - rpa_i \geq \frac{(r+\kappa)\,u_i(a_j)+\delta\sum_k\pi_k u_k(a_j)}{r+\delta+\kappa} - rpa_j \tag{57}$$

holds for every $i$ and $j$. Since (9) implies $a_i = a_j$ if and only if $a_i = a_j = 0$, (57) holds with strict inequality for any $i$ such that $a_i > 0$. Multiplying this inequality through by $\pi_i\pi_j$ and summing over all $i$ and $j$ implies $\sum_{i,j}\pi_i\pi_j\{u_i[a_i(\upsilon)] - u_i[a_j(\upsilon)]\} > 0$. The inequality is strict since for every $\upsilon$ we have $a_i > 0$ at least for $i = I$. Then, $\lim_{\upsilon\to 0}\Gamma(\upsilon) = \infty$ follows from $\eta > 0$ and the fact that

$$\lim_{\upsilon\to 0}\frac{\alpha(\upsilon)/\upsilon}{[\beta+\delta+\alpha(\upsilon)][r+\beta+\delta+(1-\eta)\alpha(\upsilon)]} = \infty.$$

Next, note that the fact that $\sum_{i,j} \pi_i \pi_j \{u_i[a_i(\upsilon)] - u_i[a_j(\upsilon)]\}$ is bounded (because $a_i(\upsilon)$ must be bounded for (15) to hold), together with

$$\lim_{\upsilon \to \infty} \frac{\alpha(\upsilon)/\upsilon}{[\beta + \delta + \alpha(\upsilon)]\,[r + \beta + \delta + (1-\eta)\alpha(\upsilon)]} = 0$$

implies that $\lim_{\upsilon \to \infty} \Gamma(\upsilon) = 0$. Finally, since $\Gamma$ is continuous, there exists some $\upsilon \in \mathbb{R}_+$ such that $\Gamma(\upsilon) = \gamma$. ■

**Proof of Proposition 6.** Substituting (39) into the planner's objective function, the problem becomes

$$\int_0^\infty \frac{\alpha+\beta}{r+\alpha+\beta} \max_{\{a_i(t)\}} \left\{ \sum_i \left[ \frac{r+\alpha+\beta}{r+\alpha+\beta+\delta} u_i[a_i(t)] + \frac{\delta}{r+\alpha+\beta+\delta} \sum_k \pi_k u_k\left[a_i(t)\right] \right] n_i(t) \right\} e^{-rt} dt$$

subject to $\sum_i n_i(t) a_i(t) \leq A$. Let

$$\mathcal{L}(t) = \sum_i \left[ \frac{r+\alpha+\beta}{r+\alpha+\beta+\delta} u_i[a_i(t)] + \frac{\delta}{r+\alpha+\beta+\delta} \sum_k \pi_k u_k[a_i(t)] \right] n_i(t) + \lambda(t)(A - \sum_i n_i(t) a_i(t)),$$

where $\lambda(t)$ is the Lagrange multiplier associated with the feasibility constraint. The planner's problem then reduces to finding, for each $t$, the sequence $\{a_i(t)\}_{i=1}^I$ that solves $\max_{\{a_i(t)\}} \mathcal{L}(t)$. The first-order necessary and sufficient condition for this problem is (42). Condition (43) is obtained by substituting the solution to the differential equation (40) into the resource constraint (41) at equality. This concludes the proof of part (a). For part (b), note that from (43), as $t \to \infty$, the shadow price of assets, $\lambda(t)$, converges to the $\lambda$ that solves

$$\sum_i \pi_i a_i^*(\lambda) = A, \tag{58}$$

where $a_i^*(\lambda)$ is the $a_i$ that satisfies (42). Comparing (58) with (15), (42) with (9), and setting $rp = \lambda$, it becomes clear that (9) coincides with (42) if and only if $\eta = 0$. ■

**Proof of Proposition 7.** The Lagrangian associated with this problem is

$$\mathcal{L} = \frac{\alpha+\beta}{\alpha+\beta+\delta} \sum_i \pi_i u_i(a_i) + \frac{\delta}{\alpha+\beta+\delta} \sum_{i,j} \pi_i \pi_j u_j(a_i) - \upsilon\gamma + \lambda \left( A - \sum_i \pi_i a_i \right),$$

where $\lambda \in \mathbb{R}_+$ is the Lagrange multiplier associated with the resource constraint $\sum_{i,j} n_{ij} a_i = A$. The first-order condition with respect to $a_i$ is

$$\frac{\alpha+\beta}{\alpha+\beta+\delta} u_i'(a_i) + \frac{\delta}{\alpha+\beta+\delta} \sum_k \pi_k u_k'(a_i) \leq \lambda, \qquad \text{“}=\text{”} \quad \text{if } a_i > 0. \tag{59}$$

As $r \to 0$, the left-hand side of (9) approaches

$$\frac{\beta + \alpha(1-\eta)}{\beta + \delta + \alpha(1-\eta)} u_i'(a_i) + \frac{\delta}{\beta + \delta + \alpha(1-\eta)} \sum_k \pi_k u_k'(a_i),$$

which coincides with the left-hand side of (59) if and only if $\eta = 0$. The first-order condition for the measure of dealers is

$$\frac{\alpha'(\upsilon)}{\beta + \delta + \alpha(\upsilon)} \sum_{i,j} \frac{\delta \pi_i \pi_j [u_i(a_i) - u_j(a_i)]}{\beta + \delta + \alpha(\upsilon)} = \gamma. \tag{60}$$

From (56) we know that, as $r \to 0$, the equilibrium condition for entry of dealers approaches

$$\frac{[\alpha(\upsilon)/\upsilon]\,\eta}{\beta + \delta + (1-\eta)\alpha(\upsilon)} \sum_{i,j} \frac{\delta \pi_i \pi_j [u_i(a_i) - u_i(a_j)]}{\beta + \delta + \alpha(\upsilon)} = \gamma,$$

which converges to (60) as $\eta \to 0$ if and only if $\alpha'(\upsilon)\upsilon/\alpha(\upsilon) = \eta$. ■