BENJAMIN LESTER
GUILLAUME ROCHETEAU
PIERRE-OLIVIER WEILL

# Competing for Order Flow in OTC Markets

We develop a model of a two-sided asset market in which trades are intermediated by dealers and are bilateral. Dealers compete to attract order flow by posting the terms at which they execute trades—which can include prices, quantities, and execution speed—and investors direct their orders toward dealers who offer the most attractive terms. We characterize the equilibrium in a general setting, and we illustrate theoretically and numerically how the model can account for several important trading patterns in over-the-counter markets, which do not emerge from existing models.

*JEL* codes: D83, G10
Keywords: OTC markets, competitive search, transaction costs.

MANY ASSETS ARE TRADED in over-the-counter (OTC) markets, including government, municipal, and corporate bonds; asset-backed securities; various types of derivatives; and currencies, to name a few. These markets have not only been growing in size, but also in economic importance; for example, the market for repurchase agreements has become crucial for the provision of liquidity, while the markets for credit default swaps and foreign exchange swaps have become essential for insuring firms and financial intermediaries against various types of shocks. Most

Pierre Mabille and Semih Uslu provided expert research assistance. We would like to thank the participants of the JMCB-SNB-Uni Bern 2013 Conference on Financial Frictions, our discussant, Julien Hugonnier, and two referees for providing comments that greatly improved the paper. Part of this project was completed when Pierre-Olivier Weill was a visiting professor at the Paris School of Economics, whose hospitality is gratefully acknowledged. The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. This paper is available free of charge at www.philadelphiafed.org/research-and-data/publications/working-papers/.

BENJAMIN LESTER *is with the Federal Reserve Bank of Philadelphia* (*E-mail:* Benjamin.Lester@phil.frb.org). GUILLAUME ROCHETEAU *is with University of California, Irvine, and Lemma at the University of Paris 2* (*E-mail:* grochete@uci.edu). PIERRE-OLIVIER WEILL *is with the University of California, Los Angeles, NBER, and CEPR* (*E-mail:* poweill@econ.ucla.edu).

Received December 13, 2013; and accepted in revised form November 21, 2014.

*Journal of Money, Credit and Banking,* Supplement to Vol. 47, No. 2 (June 2015)

often, trade in these markets is intermediated by dealers who maintain a two-sided market, simultaneously buying and selling securities on behalf of individual investors.

A recent literature, building off of the framework in Duffie, Gârleanu, and Pedersen (2005, henceforth DGP), has developed theoretical models that capture important features of intermediated OTC markets and has used these models to gain a better understanding of the factors that determine bid and ask prices, allocations, the speed of trade, and overall trading volume. These models formalize intermediated OTC markets as highly decentralized trading venues in which investors contact dealers at a constant (exogenously specified) rate, at which time they are able to trade at a price determined by *ex post* bargaining.

While this formalization has generated a number of insights, it is at odds with two important features of real-world OTC markets. First, dealers often post and commit to bid and ask prices, which are readily available to investors.[1] Second, these prices appear to play an important role in determining the speed at which different investors trade. More specifically, empirical evidence suggests that investors face a clear trade-off between the terms of trade that they are quoted and the speed at which their trade will be executed; see, for example, Hendershott and Madhavan (2015) and Li and Schürhoff (2012) for evidence from the corporate and municipal bond markets, respectively.[2] Hence, even though empirical evidence suggests that search frictions are an important element of OTC markets, it remains a challenge to formalize these frictions in a way that (i) incorporates the role of prices in directing order flow and (ii) generates the observed relationships between the prices that dealers post, the types of investors who choose those prices, and the speed at which these investors trade. In this paper, we construct and analyze a model of OTC markets to accomplish these goals.

Our framework adopts many of the basic building blocks from the literature that has evolved from DGP, and specifically the generalization by Lagos and Rocheteau (2009, henceforth LR): investors with endogenously determined asset positions periodically receive idiosyncratic shocks that affect their private valuations, at which time they attempt to rebalance their portfolios by trading with dealers in a frictional market. The major point of departure is the way we model these frictions: in contrast to the existing literature, we assume that dealers post contracts that are readily observable to

1. As Duffie (2012, p. 4) writes, "In some dealer-based OTC markets, especially those with active brokers, a selection of recently quoted or negotiated prices is revealed to a wide range of market participants. . . . In OTC markets for U.S. corporate and municipal bonds, regulators have mandated post-trade price transparency through the publication of an almost complete record of transactions shortly after they occur. . . . In OTC derivatives markets, such as the market for credit default swaps, clients of dealers can request 'dealer runs,' which are essentially lists of dealers' prospective bid and offer prices on a menu of potential trades. Dealers risk a loss of reputation if they frequently decline the opportunity to trade near these indicative prices when contacted soon after providing quotes for a dealer run."

2. In the corporate bond market, Hendershott and Madhavan (2015) find that investors sometimes choose to purchase or sell via "voice" trading because it can be quicker, despite a significant cost differential relative to trading through an auction framework. In the municipal bond market, Li and Schürhoff (2012) find a similar trade-off between execution speed and trading costs, noting that "markups increase with the matching rate of the dealers intermediating the trades." More generally, Boehmer (2005) documents that "high execution costs are systematically associated with fast execution speed, and low costs are associated with slow execution speed. This relationship holds both across markets and across order sizes."

all investors, but search frictions remain because each dealer is limited in his capacity to process multiple orders at once. This equilibrium concept is based on the model of "competitive search," first introduced by Moen (1997) in the context of the labor market, but here generalized to a two-sided market where dealers have an active role in executing the trades of both buyers and sellers on a frictionless interdealer market.

To be more precise, we assume that there is a large mass of competitive dealers posting contracts that specify a quantity that they will buy or sell on the interdealer market, along with a fee for doing so (or, equivalently, a bid–ask spread). Investors, who are heterogeneous with respect to both their asset positions and their private valuations for the asset, observe these contracts and direct their order flow to the contract of their choosing. However, dealers are capacity constrained—they cannot execute orders instantaneously, but rather orders are processed slowly, at a rate that is decreasing in the ratio of orders per dealer for this particular contract.

This formulation of asset market frictions offers several advantages. First, the process of price formation in the model more closely resembles the protocol in many OTC markets: dealers post prices, and investors send orders to these dealers with the expectation that their orders will be filled at the posted price. Second, the model is able to account for several empirical findings that do not emerge from existing models; for example, investors face the trade-off between the bid–ask spread and execution speed discussed above, and for some specifications, these bid–ask spreads decrease with the size of the trade, which is consistent with evidence from certain OTC markets. Third, the model generates new normative implications as well; for example, in a simple extension of the benchmark model, we show that welfare is unambiguously increasing in the fraction of investors who can observe posted contracts and direct their search, thus highlighting a margin through which transparency in financial markets can improve allocations. Finally, the competitive search setting offers a number of technical advantages, too: it avoids several inefficiencies that arise from the Nash bargaining protocol, such as the well-known holdup problem (see, e.g., Hosios 1990); it can easily accommodate the introduction of private information (see, e.g., Guerrieri, Shimer, and Wright 2010), and it also offers a more tractable environment for the analysis of aggregate shocks (see, e.g., Menzio and Shi 2011). Hence, we believe the model provides a useful benchmark for future modifications and extensions.

The structure of the paper is as follows. We begin with the most general environment, where investors are free to hold any (positive) asset position, and they periodically receive idiosyncratic shocks to their private valuations that are drawn from an arbitrary distribution with finite support. We provide a full characterization of steady-state equilibria, establish existence, and flesh out a number of properties of equilibrium prices, allocations, and trading speeds. We show that dealers design contracts to respond to investors' trading needs, and investors subsequently sort themselves across posted contracts according to their private valuations and current asset holdings. In particular, the investors whose preferences are most severely misaligned with their asset holdings choose to trade with dealers who offer fast execution times but high trading fees.

Then, in order to gain additional insights and to compare our model more squarely with the existing literature, we analyze two special cases in detail. First, we consider an environment in which agents have one of two private valuations and their asset holdings lie in the set {0, 1}, as in DGP. These restrictions make each investor's choice of asset holdings trivial and allow us to focus on the asset prices, trading fees, and execution times that prevail in the unique equilibrium. We show that the equilibrium can be solved in closed form when search frictions are small, as in Vayanos and Weill (2008) and Weill (2008), which allows for a large set of comparative statics. Interestingly, we establish that the relationship between intermediation fees and the sellers' share of the surplus (or implicit bargaining power) is hump-shaped, whereas bid–ask spreads in DGP are increasing in the dealers' exogenous bargaining power. We also show that the price discount, which indicates the rate at which the asset price falls relative to the Walrasian benchmark due to search frictions, is nonmonotonic with the dealers' bargaining power.

For the second special case, we return to the general model with asset holdings in $\mathbb{R}_+$ and an arbitrary number of preference shocks, but we adopt a Leontief specification for the expected execution time; that is, we assume there is strict complementarity between the measure of dealers and the measure of orders to be executed. This trading technology generates average execution delays that are constant across active submarkets, allowing us to focus on the implications of our equilibrium concept for investors' endogenous asset holdings. We characterize the equilibrium when the dealers' entry cost is small, which serves two purposes. First, the analysis is tractable in this region of the parameter space, and hence we can study how asset demand, bid–ask spreads, and trading volume respond to changes in the economic environment. Second, the equilibrium asset holdings and the interdealer prices are the same as in LR when, in their model, the exogenous bargaining power of dealers is equal to zero. Therefore, this class of equilibria is amenable to a direct comparison with the existing literature. Our analysis reveals that the predictions of the competitive search model differ markedly from the earlier literature in one dimension: the relationship between trading costs and trade sizes. In earlier models with random matching and *ex post* bargaining, such as that of LR, the trading cost per unit traded increases with the size of the trade. In contrast, under competitive search, per unit trading costs decrease with trade sizes.[3] This prediction of our model is in line with the evidence documented by Schultz (2001), Edwards, Harris, and Piwowar (2007), Green, Hollifield, and Schürhoff (2007), and Hollifield, Neklyudov, and Spatt (2012) for OTC markets. Therefore, the structure of the market helps explain the empirical relationship between trading costs and trade sizes.[4]

3. The reason is that dealers reap a constant fraction of the surplus in LR, and the surplus is convex in trade size. In our competitive search model with Leontief technology, by contrast, the price-setting mechanism implies that investors pay a fixed trading cost to trade any quantity.

4. See Zhang (2012) for a different explanation in a model with *ex post* bargaining under asymmetric information.

Finally, in Section 4, we develop a numerical example within the context of our general model. The purpose is twofold. First, this exercise illustrates that our model can generate quantitative outcomes of reasonable magnitude, in terms of trade size, execution speed, and trading cost. Second, we use this calibrated model to perform comparative statics exercises. In particular, we study how the trading technology, which determines the time it takes for dealers to execute orders, affects the *average* per unit trading cost that investors pay, as well as the *elasticity* of this per unit trading cost to trade size, in the cross section. This analysis suggests that cross-sectional data on transaction costs could be informative about some important features of the trading technology.

*Related Literature*

Our paper is part of a long tradition in the market microstructure literature, starting at least with Ho and Stoll (1983), studying competition between dealers who serve outside customers. The description of the asset market that involves search and pairwise meetings between dealers and investors, where dealers have access to a competitive interdealer market, is based on DGP and LR. Other papers in this literature include Weill (2007), Gavazza (2011), and Lagos, Rocheteau, and Weill (2011). There are also other search-theoretic models of financial intermediation with price posting. Spulber (1996) describes an environment with competing price-setting dealers. However, in contrast to our analysis, the contract posted by dealers (bid and ask prices) can be observed only after a time-consuming search.[5] For a more thorough literature review of the search-theoretic approach to OTC markets, see Rocheteau and Weill (2011).

Search-and-matching models offer a natural platform to study the interactions of investors trading at different speed. Some papers in the literature study asset pricing when speed differences are exogenously given, for example, Feldhütter (2012) and Neklyudov (2012). Others are explicit about the trade-off between execution speed and trading costs.

An early contribution is the sequential search model of block trading by Burdett and O'Hara (1987). More recent work includes the models of Melin (2012) and Praz (2012), in which investors can trade simultaneously in a Walrasian and an OTC market, or that of Pagnotta and Philippon (2012), in which investors can choose between competing trading platforms that offer different execution speeds. We study the choice of execution speed using a different approach, based on the notion of competitive search, initially developed by Moen (1997). Mortensen and Wright (2002) and Sattinger (2003) interpret this equilibrium notion as one where competing brokers or market-makers set up markets and charge entry fees to participants. However, in contrast to our model, the brokers in these models do not contribute to the matching

5. Hall and Rust (2003) extend Spulber (1996) by introducing a second type of middlemen called market-makers. Each market-maker posts publicly observable prices while posted prices of middlemen are not observable.

of orders on both sides of the market and do not intermediate trades. Weill (2007, appendix IV) offers some preliminary analysis of a competitive search equilibrium without free entry of dealers. Rocheteau and Weill (2011, p. 272) apply competitive search to a simple model of an OTC market where trades are not intermediated by dealers and asset holdings are restricted to {0, 1}.

Competitive search has also been recently applied to asset markets by Guerrieri and Shimer (2012, Forthcoming) and Chang (2012) in order to study the impact of asymmetric information (about both common and private values) on liquidity in OTC markets. To do so, these authors build on Guerrieri, Shimer, and Wright (2010) and consider "one-sided" competitive search models with {0, 1} asset holdings, in which buyers post contracts to attract privately informed sellers; also see Inderst and Müller (2002), who study a market with durable goods under adverse selection and competitive search. While we deal only with asymmetric information about private values (see also Faig and Jerez 2006), we explicitly characterize an equilibrium in a two-sided market, where prices are posted by dealers who simultaneously buy and sell assets from investors, and we are able to remove asset-holding restrictions.[6] Finally, competitive search has been applied to markets for real assets, such as housing, by Albrecht, Gautier, and Vroman (2013), Diaz and Jerez (2013), Lester, Visschers, and Wolthoff (2013), and Stacey (2012). These models do not have dealers making two-sided markets.

## 1. ENVIRONMENT

Time is continuous and goes on forever. The economy is populated with two types of infinitely lived agents: a unit measure of investors and a large measure of dealers. Both types of agents discount the future at rate $r > 0$. There is one long-lived asset in fixed supply $A \in \mathbb{R}_+$. There is also a perishable good, the numéraire, which is produced and consumed by all agents.

### *1.1 Preferences*

The instantaneous utility function of an investor is $u_i(a) + c$, where $a \in \mathbb{R}_+$ denotes the investor's asset holdings, $c \in \mathbb{R}$ denotes the net consumption of the numéraire good (with $c < 0$ if the investor produces more than he consumes), and $i \in \{1, ..., I\} \equiv \mathcal{I}$ indexes an investor's type, where $1 < I < \infty$. [7] We assume

6. Watanabe (2013) also utilizes a model of competitive search to analyze the role of middlemen in asset markets. In his model, middlemen are assumed to have a greater capacity to store inventory than ordinary sellers. In contrast to our assumption of a two-sided asset market, he assumes that middlemen can acquire these inventories—that is, they can purchase assets from sellers—in a frictionless market, before posting prices and selling to buyers in a market with search frictions.

7. The specification for the utility function is borrowed from DGP, where $a \in \{0, 1\}$, and LR, where $a \in \mathbb{R}_+$. In the case of a real asset, such as houses or physical capital, $u_i(a)$ is a standard utility or production function. In the case of a financial asset, $u_i$ is a reduced-form utility function that captures the

that $u_i(a)$ is continuously differentiable, with $u_i'(a) > 0$, $u_i''(a) < 0$, $u_i'(0) = \infty$, and $u_i'(\infty) = 0$ for all $i \in \mathcal{I}$. We also assume that $u_i'(a) < u_{i+1}'(a)$ for all $a > 0$ and $i < I - 1$, so that investors with higher $i$ have higher demand for the asset. This condition holds, for example, for the multiplicative specification in LR, where $u_i(a) \equiv \varepsilon_i v(a)$ for $\varepsilon_1 < \varepsilon_2 \ldots < \varepsilon_I$.

An investor's preferences change over time according to a Poisson process with arrival rate $\delta$. Conditional on receiving a preference shock, an investor of type $i$ draws a new type $j \in \mathcal{I}$ with probability $\pi_{ij}$, where $\sum_{j \in \mathcal{I}} \pi_{ij} = 1$ for all $i \in \mathcal{I}$. These type-switching processes are assumed to be independent across investors. Unlike investors, dealers neither receive utility flow from holding an asset, nor can they hold asset inventories.[8] Their instantaneous utility is simply $c$, the net consumption of the numéraire good.

### *1.2 Trade*

All trades are bilateral and are intermediated by dealers; that is, they involve one dealer and one investor. Dealers have continuous access to a competitive interdealer market in which they can trade on behalf of investors.[9] In order to attract orders from investors, dealers post (and commit to) a publicly observable contract $\sigma = (q, \phi)$ specifying that the dealer will trade a quantity $q$ at the prevailing interdealer market price $p$ in exchange for an intermediation fee $\phi$.[10] If $q > 0$, the dealer buys the specified amount and delivers it to the investor. If $q < 0$, the dealer acquires the asset from the investor and immediately resells it on the interdealer market. There is free entry into the dealer market: dealers can choose to enter by posting a contract at a

various reasons why investors may want to hold different quantities of the asset, such as differences in liquidity or hedging needs. The quasi-linear specification, which is common to both monetary and labor search models, is adopted for tractability. Alternatively, Duffie, Gârleanu, and Pedersen (2007), Vayanos and Weill (2008), and Gârleanu (2009) adopt constant absolute risk aversion (CARA) preferences. One could also generalize preferences along the lines suggested by Wong (2013) in the context of a monetary search model.

8. The assumption that dealers cannot hold assets is without loss of generality when analyzing steady-state equilibria, as we do in this paper. See Weill (2007) and Lagos, Rocheteau, and Weill (2011) for dynamic equilibria where dealers hold positions.

9. It should be emphasized that our dealers are different from the market-makers in Moen (1997) and Mortensen and Wright (2002). In particular, the market-makers in these papers are passive third parties who announce prices in each submarket; sellers then make an entry decision, buyers search, and the two parties trade directly. In our model, market-makers play a more active role: they decide whether to enter the market, then trade with buyers and sellers, and finally settle these trades on the interdealer market. It is in this sense that we generalize the competitive search framework to a "two-sided market." The assumption that the interdealer market is frictionless, which was first introduced by DGP, is a simplification that allows us to focus on the interaction between investors and dealers. In reality, interdealer markets may also be frictional; these additional frictions could be modeled, though the analysis would be more cumbersome.

10. Alternatively, one could let dealers post a menu, $\phi(q)$, that specifies the intermediation fee as a function of the quantity of assets traded. Anticipating that a submarket will open for each type of investor, it is without loss in generality that we restrict $\sigma$ to a flat fee contract. For an elaboration on this point in the context of a monetary search model, see appendix B in Lagos and Rocheteau (2004). Also, an alternative interpretation of the model is one with large representative dealer firms composed of a continuum of individual dealers who can be allocated to different submarkets. In this case, the relevant schedule is a list of triples, $(q, \phi, \theta)$.

flow cost $\gamma > 0$, which captures the ongoing costs of advertising their services to investors, maintaining access to the interdealer market, and so on.[11]

On the other side of the market, investors observe the contracts that have been posted and send their order to one of these contracts. Dealers, however, cannot process these orders instantly. In reality, there are a variety of factors that can delay either the process by which dealers receive orders or the process by which they execute them on the interdealer market. These factors include: technological constraints on the ability of computers or communication lines to process orders; institutional constraints on a dealer's ability to execute, clear, settle, and confirm that orders have been off-loaded in the interdealer market; cognitive constraints on a dealer's ability to pay attention to multiple orders at once; and potentially even delays caused by information asymmetries, such as the requirement that a dealer verifies the quality or value of collateral that has been posted. In this paper, we do not take a precise stand on the reason why dealers take time to process an order. Instead, we assume that orders are executed according to some abstract order-processing technology, at a rate that is decreasing in the ratio of orders per dealer for this particular contract.[12]

Formally, suppose that there is a measure $d$ of dealers posting a particular contract, and a measure $o$ of orders for this contract. We assume that, at each instant, the flow of orders processed and executed by dealers is equal to $m(d, o)$, a function that has constant returns to scale, is strictly increasing, strictly concave, and twice continuously differentiable with respect to its two arguments, and satisfies Inada conditions.[13] Assuming further that each order is equally likely to be executed at any point in time, the time it takes to execute any particular order for this contract is an exponentially distributed random variable with parameter $\alpha(\theta) \equiv m(d, o)/o = m(\theta, 1)$, where $\theta \equiv d/o$.[14] Symmetrically, a dealer who has posted this contract executes it at an exponentially distributed time with intensity $\alpha(\theta)/\theta$. These assumptions on $m(d, o)$ imply that $\alpha(\cdot)$ is continuous, strictly increasing, strictly concave, and satisfies $\alpha(0) = 0$, $\alpha(\infty) = \infty$, and $\alpha(\infty)/\infty = 0$. Moreover, the strict concavity of $\alpha(\cdot)$ implies that $\alpha(\theta)/\theta$ is strictly decreasing in $\theta$. Finally, if an investor's idiosyncratic

11. Our free-entry of dealers is similar to the one in LR. Similar free-entry conditions have been used routinely in labor search models (Pissarides 2000), in monetary search models (Rocheteau and Wright 2005), and in search models with middlemen (Rubinstein and Wolinsky 1987, Shevchenko 2004). Following Moen (1997), free-entry conditions are also typically assumed in competitive search environments as the entry cost pins down the value of agents on one side of the market, but one could also consider a competitive search environment with fixed populations.

12. Of course, depending on the precise origin of these frictions, the ratio of orders to dealers might not be sufficient statistic for execution delays. For example, if the relevant constraints lie in a dealer's ability to trade each share on the interdealer market, then the expected delay could potentially depend on the *size* of the order—that is, larger orders could take longer to fill—and hence the *distribution* of quantities that a dealer is asked to fill could be relevant. We leave these types of extensions, and other modifications to the matching function, for future work.

13. Whether or not the order execution technology has constant returns to scale remains an open question in the context of asset markets. Unfortunately, in contrast to the labor market—where reliable data exist for the number of unemployed workers, the number of vacancies, and the number of matches that form in a particular labor market—the analogous data for financial markets is elusive.

14. Burdett, Shi, and Wright (2001) formally illustrate how capacity constraints can generate a function $\alpha(\theta)$ with the properties we assume. This formulation of meeting frictions is also consistent with the "telephone line" Poisson queueing processes described in Stevens (2007) and Sattinger (2010).

valuation changes before a standing order is executed, we assume that he can cancel the unfilled order at no cost and redirect a new order toward a different contract.

## 2. EQUILIBRIUM

In this section, we define and derive conditions for a steady-state equilibrium. These conditions describe (i) the optimal entry and posting behavior of dealers, taking as given the order flow they receive from posting any contract and the interdealer price; (ii) the optimal order submission strategy of investors, taking as given prices and the set of contracts that are available to them; and (iii) the market tightness in each active submarket that is consistent with both a stationary distribution across investor types and market clearing.

### *2.1 Definition*

Let $\Sigma$ denote the set of all possible contracts and $\Sigma^\star$ the set of contracts offered in equilibrium. We assume that an investor can always choose not to send any order, so that $\mathbf{0} = (0, 0) \in \Sigma^\star$. We define a submarket as a contract, $\sigma \in \Sigma$, a set of dealers posting this contract, and a set of investors sending an order for this contract. The dealer-to-investor ratio, or market tightness, prevailing in a submarket for contract $\sigma \in \Sigma$ is denoted by $\Theta(\sigma)$. The function $\Theta(\sigma)$ is defined for all possible contracts, not only those offered in equilibrium. The set of all possible investors' types, $(i, a)$, is $\mathcal{N} \equiv \mathcal{I} \times \mathcal{A}$. For simplicity, we let $\mathcal{A} = \mathbb{R}_+$ but our formulation and proofs remain identical if there are asset-holding restrictions. We denote the support of the equilibrium stationary distribution of investors' types by $\mathcal{N}^\star \subseteq \mathcal{N}$.

*Dealers.* The profits of a dealer who posts a contract $\sigma \in \Sigma$ are denoted by $\Pi(\sigma)$ and solve:

$$r\Pi(\sigma) = -\gamma + \frac{\alpha\left[\Theta\left(\sigma\right)\right]}{\Theta\left(\sigma\right)}\phi.$$

The first term on the right-hand side is the flow cost of posting a contract, while the second term is the expected fee received by a dealer, that is, the Poisson rate at which a dealer receives the order, $\alpha\left[\Theta(\sigma)\right]/\Theta(\sigma)$, times the fee paid by the investor upon execution of the order, $\phi$. Therefore, the zero-profit condition of dealers can be written as

$$\Pi(\sigma) \leq 0, \text{ with equality if } \sigma \in \Sigma^\star \text{ and } \sigma \neq \mathbf{0}. \tag{1}$$

The expected profits of a dealer are zero in any active submarket, $\sigma \in \Sigma^\star$, and they are nonpositive in inactive submarkets, $\sigma \in \Sigma \backslash \Sigma^\star$, since otherwise posting a contract in that submarket would be profitable.

*Investors.* Let $V_i(a, \sigma, \theta)$ denote the expected lifetime payoff of an investor of type $(i, a) \in \mathcal{N}^\star$ who sends an order for contract $\sigma = (q, \phi)$ given market tightness $\theta$:

$$V_i(a, \sigma, \theta) = \frac{u_i(a) + \delta \sum_{j \in \mathcal{I}} \pi_{ij} V_j^\star(a) + \alpha(\theta) \left[ V_i^\star(a + q) - pq - \phi \right]}{r + \delta + \alpha(\theta)}. \quad (2)$$

The first term in the numerator is the investor's utility flow; the second term is the investor's expected continuation payoff, conditional on switching types, which occurs with Poisson intensity $\delta$; and the third term is the continuation payoff when his order is executed, which occurs with Poisson intensity $\alpha(\theta)$. The denominator is the effective discount rate. This investor's maximum attainable utility, which we denote by $V_i^\star(a)$, is defined as

$$V_i^\star(a) = \sup_{\sigma \in \Sigma^\star} V_i \left[ a, \sigma, \Theta(\sigma) \right].$$

*Market tightness.* We adopt the convention that $\Theta(\mathbf{0}) = 0$. For $\sigma \neq \mathbf{0}$, we assume that

$$\Theta(\sigma) = \inf \left\{ \theta \geq 0 \; : \; V_i(a, \sigma, \theta) > V_i^\star(a) \text{ for some } (i, a) \in \mathcal{N}^\star \right\}, \quad (3)$$

and $\Theta(\sigma) = \infty$ if this set is empty. This definition captures the idea that if a dealer posts a contract $\sigma$, then the investors who value it most will direct their order flow to this contract until they are indifferent between this contract and their best alternative. Put differently, if the market tightness for contract $\sigma$ were greater than $\Theta(\sigma)$, then there would be a positive measure of investors who would benefit strictly from this contract. Orders sent by these investors would further reduce tightness until it is equal to $\Theta(\sigma)$.[15]

*Market clearing and the distribution of investors' states.* Let $n_i(da)$ denote the steady-state measure of investors over the set of types, $\mathcal{N}$. The measure $n_i(da)$ must satisfy the following two identities:

$$\sum_{i \in \mathcal{I}} \int_{\mathcal{A}} n_i(da) = 1, \quad (4)$$

$$\sum_{i \in \mathcal{I}} \int_{\mathcal{A}} a n_i(da) = A. \quad (5)$$

15. Note that we are taking the infimum over all $\theta$ that create a *strict* utility improvement. This is important because otherwise $\Theta(\sigma) = 0$ for all $\sigma$: those investors who have no gains from trade, $V_i^\star(a) = V_i(a, \mathbf{0}, 0)$, would flow into the submarket in the expectation that they will not be able to trade anyway, and thus $\theta = 0$.

Equation (4) imposes that the measures of investors add up to one. Equation (5) ensures that investors hold the entire asset supply and that the interdealer market clears.

The order submission strategy of an investor of type $(i, a)$ can be represented by a probability measure, $\zeta(d\sigma \mid i, a)$, over some support

$$\Sigma_i^\star(a) \subseteq \arg\max_{\sigma\in\Sigma^\star} V_i\left[a, \sigma, \Theta(\sigma)\right].$$

Then, in a steady-state equilibrium, $n_i(da)$ must satisfy

$$\delta \sum_{j\in\mathcal{I}} n_j(da)\pi_{ji} + \int_{\mathcal{A}} \int_{\Sigma_i^\star(a')} \zeta(d\sigma \mid i, a')\alpha[\Theta(\sigma)]\mathbb{I}_{\{a'+q=a\}}$$
$$= \delta n_i(da) + n_i(da) \int_{\Sigma_i^\star(a)} \zeta(d\sigma \mid i, a)\alpha[\Theta(\sigma)]. \qquad (6)$$

The left-hand side of (6) represents the inflow of investors, while the right-hand side represents the outflow. On both sides the first term represents the flow due to type switching, and the second term is the flow due to trade.

Definition 1. *A steady-state competitive search equilibrium is a list composed of an interdealer market price, p; a set of open submarkets, $\Sigma^\star$; a market tightness function, $\Theta(\sigma)$; a collection of value functions, $V_i^\star(a)$; an order submission strategy $\zeta(d\sigma \mid i, a)$ with support $\Sigma_i^\star(a)$; and a measure on the set of investors' types, $n_i(da)$, with support $\mathcal{N}^\star \subseteq \mathcal{A} \times \mathcal{I}$, satisfying (1)–(6).*

### 2.2 *Characterization*

As is typically true in competitive search models, there is a dual formulation of the problem according to which an investor's expected utility is maximized with respect to trade size, $q$; intermediation fee, $\phi$; and market tightness, $\theta$, subject to dealers' zero-profit condition, $[\alpha(\theta)/\theta]\phi = \gamma$. This dual problem can be represented conveniently by the following flow Bellman equation:

$$rV_i^\star(a, p) = u_i(a) + \delta \sum_{j\in\mathcal{I}} \pi_{ij} \left[V_j^\star(a, p) - V_i^\star(a, p)\right]$$
$$+ \max_{q\geq -a,\theta,\phi} \left\{\alpha(\theta)\left[V_i^\star(a+q, p) - pq - \phi - V_i^\star(a, p)\right]\right\}$$
$$\text{subject to } \frac{\alpha(\theta)}{\theta}\phi = \gamma.$$

If a utility-maximizing contract, $(q, \phi, \theta)$, was not offered in a candidate equilibrium, a dealer would have a profitable deviation by offering the same contract with a slightly higher fee, thereby attracting the type of investors for whom this contract is optimal.

Substituting $\phi = \theta\gamma/\alpha(\theta)$ from the dealer zero-profit condition into the Bellman equation, we obtain the following characterization for investors' value functions.

Proposition 1. *In any equilibrium, for all $(i, a) \in \mathcal{N}^\star$, investors' value functions solve*

$$V_i^\star(a, p) = \max_{q \geq -a,\, \theta} \frac{u_i(a) + \delta \sum_{j\in\mathcal{I}} \pi_{ij} V_j^\star(a, p) + \alpha(\theta)\left[V_i^\star(a+q, p) - pq\right] - \gamma\theta}{r + \delta + \alpha(\theta)}. \tag{7}$$

*Moreover, there exists a unique collection $\{V_i^\star(a, p)\}_{i\in\mathcal{I}}$, defined over $S = \mathcal{I} \times \mathcal{A} \times (0, \infty)$, that satisfy (7). Each $V_i^\star(a, p)$ is continuous in $(a, p)$ and strictly increasing in $a$.*

The proposition offers an alternative representation of the Bellman equations, which is useful for at least two reasons. First, it allows us to solve for investors' values at each point in the support of the distribution of investors' types, $\mathcal{N}^\star$, without having to characterize $\Sigma^\star$, the set of contracts offered in equilibrium. Second, it shows that our model is observationally equivalent to one in which investors directly choose their order-execution intensity, $\mu \equiv \alpha(\theta)$, but have to incur a convex cost $\gamma\alpha^{-1}(\mu)$. Put differently, the outcome of competition and free entry is that dealers act "as if" they knew investors' private utility type and were making an optimal search intensity decision on their behalf.[16]

Next, we proceed to establish the existence of an equilibrium. First, note from the right-hand side of (7) that investors' post trade asset holdings belong to $\arg\max_{a'}\{V_i(a', p) - pa'\}$, so they are independent of $a$, the investor's current asset holdings.[17] For simplicity, let us assume for now that this program has a unique maximizer, which we denote by $a_i$.[18] We look for a stationary equilibrium in which the support of the distribution of types is $\mathcal{I} \times \{a_1, \ldots, a_I\}$. This is intuitive since, at his first opportunity to trade, an investor chooses asset holdings in the support $\{a_1, \ldots, a_I\}$, and then continues to keep his holdings in this support forever after. Let $\theta_i(a_j)$ be the maximizer of the Bellman equation, (7), when the investor's type is $(i, a_j)$. The inflow–outflow equations for the steady-state distribution, $n_i(a_j)$, are

$$0 = \delta \sum_{k\in\mathcal{I}} n_k(a_j)\pi_{ki} - \delta n_i(a_j) - \alpha[\theta_i(a_j)]n_i(a_j), \quad \text{if } i \neq j; \tag{8}$$

16. Note, however, that the present competitive search model is not equivalent to a random search model with endogenous search intensities and bargaining. Indeed, under *ex post* bargaining, the choice of search intensities generates externalities that are not internalized by the pricing mechanism. In contrast, the outcome of a model with endogenous search intensities but competitive price posting would typically be constrained efficient.

17. As in LR this property of the equilibrium follows from the quasi-linear specification for investors' utility function. If zero is an optimal market tightness in (7), then optimal asset holdings can take any value on the real line, so they do not depend on $a$ either.

18. Our proof in the Appendix deals with the general case in which there may be multiple maximizers.

$$0 = \delta \sum_{k \in \mathcal{I}} n_k(a_i)\pi_{ki} - \delta n_i(a_j) + \sum_{k \in \mathcal{I}} \alpha[\theta_i(a_k)] n_i(a_k), \quad \text{if } i = j; \tag{9}$$

$$1 = \sum_{(i,j) \in \in \mathcal{I}^2} n_i(a_j). \tag{10}$$

Since finite-state Markov chains have at least one ergodic distribution, it follows that this system of equations has at least one solution (see Lemma A2 in the Appendix). Given any solution, we define the aggregate demand as $\mathcal{D}(p) = \sum_{i,j} n_i(a_j)a_j$, and we look for a market-clearing price, that is, some $p \in \mathbb{R}_+$ such that $A \in \mathcal{D}(p)$.

PROPOSITION 2. *A stationary equilibrium exists.*

The first part of the proof uses the Intermediate Value Theorem to establish that there exists a market-clearing price. Having found a candidate market-clearing price, the second part of the proof constructs the remaining equilibrium objects in such a way that all conditions of Definition 1 are satisfied. The construction goes as follows. We first obtain the equilibrium price by solving $A \in \mathcal{D}(p)$. We then obtain investors' optimal asset holdings, $\{a_i\}_{i \in \mathcal{I}}$, the market tightness in each submarket, $\{\theta_i(a_j)\}_{(i,j) \in \mathcal{I}^2}$, and the distribution of types, $\{n_i(a_j)\}_{(i,j) \in \mathcal{I}^2}$. From these conditions, along with the free-entry condition, (1), we obtain trade sizes and intermediation fees:

$$q_i(a_j) = a_i - a_j, \tag{11}$$

$$\phi_i(a_j) = \gamma \frac{\theta_i(a_j)}{\alpha[\theta_i(a_j)]}. \tag{12}$$

The order submission strategy of an investor in state $(i, a_j)$ is to direct his order flow toward the submarket $\sigma_i(a_j)$.[19] The set $\Sigma^\star$ is then composed of all $\sigma_i(a_j)$. The value functions are defined as the solutions to the Bellman equations:

$$V_i^\star(a) = \max_{k,\ell \in \mathcal{I}^2} V_i\left[a, \sigma_k(a_\ell), \theta_k(a_\ell)\right].$$

By construction, for $(i, a_j) \in \mathcal{N}^\star$, $V_i^\star(a_j)$ coincides with the solution of our auxiliary Bellman equation (7) evaluated at $(i, a_j)$. Finally, we let the market tightness associated with any contract $\sigma \neq \mathbf{0}$ be defined as in equation (3).

To conclude this section, we recapitulate some key properties of the competitive search equilibrium of an OTC market. First, from the free-entry condition of dealers, in any active submarket

$$\frac{\alpha(\theta)}{\theta}\phi = \gamma.$$

19. Note that the equilibrium price equates aggregate demand and supply, but does not explicitly equate "buy" and "sell" order flows in the interdealer market. As we show in Appendix A.3, this latter condition is actually implied by the steady-state conditions (8) and (9).

From the dual formulation of a competitive search equilibrium, this condition implies that investors face a trade-off between the cost at which they can readjust their asset holdings, $\phi$, and the order execution time, $1/\alpha(\theta)$.[20] As suggested by Boehmer (2005) and others, this trade-off is relevant in practice across trading venues. Second, from the first-order condition of the dual problem and the free-entry condition of dealers,

$$\frac{\theta_i(a_j)\alpha'[\theta_i(a_j)]}{\alpha[\theta_i(a_j)]}\left[V_i^\star(a_i,p)-p(a_i-a_j)-V_i^\star(a_j,p)\right]=\frac{\gamma\theta_i(a_j)}{\alpha[\theta_i(a_j)]}=\phi_i(a_j).$$

The dealer's fee is a fraction $\theta\alpha'(\theta)/\alpha(\theta)$ of the match surplus, where this fraction is equal to the elasticity of the matching function. This corresponds to the Hosios (1990) condition for efficiency in markets with search frictions. Relative to models with *ex post* bargaining, such as those of DGP and LR, where dealers' bargaining power is exogenous, here the dealers' share of the match surplus is endogenous and equal to the contribution of dealers to the matching process. This means that the externalities associated with dealers' entry decisions are internalized through the pricing mechanism. An additional consequence of the competitive search formulation is that investors are not subject to a holdup problem when choosing asset holdings; in particular, for a given $\theta$, competition between dealers makes it "as if" investors have an unintermediated access to the market. So, in terms of policy implications, our model suggests that more transparent OTC markets, in the sense of contracts being publicly announced and search being directed, have better welfare properties than opaque markets with purely random search and *ex post* bargaining. We come back to this point in greater detail in the next section. Third, the number of active submarkets will reflect the endogenous heterogeneity across investors since, according to the dual formulation, each investor opens a submarket that is optimal given his state $(i, a_j)$ with $j \neq i$. Hence, our model offers a rationalization for a high degree of market segmentation.[21]

## 3. LIQUIDITY AND PRICES IN TWO SPECIAL CASES

In the remainder of the paper we focus on two special cases: (i) the case where asset holdings are restricted to $\mathcal{A} = \{0, 1\}$ and the set of investors' private valuations is $\mathcal{I} = \{\ell, h\}$, but the matching function is general, and (ii) the case where asset

20. If $\alpha$ was a function of the size of the trade, $q$, then the model would predict a more general trade-off between $\phi$, $q$, and $\theta$. Such an extension of the matching technology in asset markets is left for future research. Pagnotta and Philippon (2012) also consider a model based on LR where investors face a trade-off between entry fees and speed. In their model, market-makers set up platforms where investors can access a competitive asset market at a given Poisson rate in exchange for a fee. In contrast to our model, individual trades are not intermediated and the extent of market segmentation is restricted.

21. Pagnotta and Philippon (2012) argue that "a major feature of the new trading landscape is fragmentation," where market fragmentation corresponds to the phenomenon according to which securities are now traded in multiple markets with different characteristics in terms of quality of execution and trading costs.

holdings are unrestricted in $\mathcal{A} = \mathbb{R}_+$ and the set of private valuations is $\mathcal{I} \subset \mathbb{N}$, but the matching function is Leontief. The first case will allow us to focus on order flows, speeds of execution, and trading costs across markets, taking as given asset positions. The second case will generate an endogenous distribution of asset holdings, allowing us to focus on relationship between trade size and trading cost.

### *3.1 Order Flows, Speeds of Execution, and Trading Costs*

In this section, we consider a setting with restricted asset holdings, $\mathcal{A} = \{0, 1\}$, and two utility types, $\mathcal{I} = \{\ell, h\}$, with respective utility flows for the asset, $u_\ell < u_h$.[22] Without loss of generality, we assume that $\pi_{\ell,h} = \pi_h$ and $\pi_{h,\ell} = \pi_\ell$. First, we will provide a characterization of equilibrium objects and show existence and uniqueness. Then, we derive analytical comparative statics for the limit economy with small search frictions. Finally, for large frictions, we offer comparative statics by way of a numerical example.

*Bellman equations.* Let $\Delta V_i \equiv V_i(1) - V_i(0)$ denote the reservation value of an investor of type $i \in \{\ell, h\}$, that is, the expected utility of owning the asset minus the expected utility of being without the asset. From Proposition 1, $\Delta V_h$ solves the following flow Bellman equation:

$$r\Delta V_h = u_h + \delta\pi_\ell\,(\Delta V_\ell - \Delta V_h) + \max_{\theta\geq 0}\{\alpha(\theta)\,(p - \Delta V_h) - \gamma\theta\} - \max_{\theta\geq 0}\{\alpha(\theta)\,(\Delta V_h - p) - \gamma\theta\}. \tag{13}$$

The right-hand side of (13) decomposes the flow reservation value into four terms. The first term, $u_h$, is the flow value of owning the asset. The second term captures the fact that with intensity $\delta\pi_\ell$, a high-valuation investor switches to a low type, in which case his reservation value drops from $\Delta V_h$ to $\Delta V_\ell$. The third and fourth terms are, respectively, the flow values of searching to sell and buy. The flow value of searching to sell enters the Bellman equation with a positive sign because this option is available to the investor only if he already owns the asset. Therefore, it raises the net utility of owning an asset. By contrast, the flow value of searching to buy enters with a negative sign because this option is only available to investors who do not own the asset.

Similarly, the Bellman equation for $\Delta V_\ell$ is

$$r\Delta V_\ell = u_\ell + \delta\pi_h\,(\Delta V_h - \Delta V_\ell) + \max_{\theta\geq 0}\{\alpha(\theta)\,(p - \Delta V_\ell) - \gamma\theta\} - \max_{\theta\geq 0}\{\alpha(\theta)\,(\Delta V_\ell - p) - \gamma\theta\}. \tag{14}$$

22. In order to extend our existence proof to the case with restricted asset holdings, we make two additional assumptions on the finite grid $\mathcal{A} \subset \mathbb{R}_+$. First, to ensure that investors can hold all the asset supply, we assume that $\max \mathcal{A} > A$. Second, to keep prices bounded we assume that $u_1(\min \mathcal{A})$ is bounded away from $-\infty$ .

LEMMA 1. *In any equilibrium* $\Delta V_h > \Delta V_\ell$ *and* $p \in (\Delta V_\ell, \Delta V_h)$.

The first result of Lemma 1 states that high-valuation investors have strictly higher reservation values than low-valuation investors, which follows from the assumption that their utility flow from holding the asset is strictly larger. The second result states that the equilibrium price must lie between the reservation value of high and low types, since otherwise the market would not clear. A consequence of these two results is that, in equilibrium, low-valuation investors have strict incentives to sell and high-valuation investors have strict incentives to buy.

The properties reported in Lemma 1 also hold in DGP, except for one key difference. In DGP, the interdealer market price is, generically, at a corner, $p \in \{\Delta V_\ell, \Delta V_h\}$. This result arises because buyers and sellers contact dealers at the same rate, so that $p$ must adjust to make investors on the long side of the market indifferent between trading and not. If $\pi_h > A$, then buyers are on the long side and $p = \Delta V_h$. Conversely, if $\pi_h < A$, then sellers are on the long side and $p = \Delta V_\ell$. Such prices, $\{\Delta V_\ell, \Delta V_h\}$, cannot be the basis of an equilibrium under competition for order flows: indeed, if investors on one side of the market were indifferent between trading or not at the interdealer price, that is, $p = \Delta V_i$ for some $i \in \{\ell, h\}$, then these investors would not be willing to pay any intermediation fee and dealers would have no incentives to make a market for such investors. Therefore, under competitive search the price has to lie strictly between $\Delta V_\ell$ and $\Delta V_h$, so that dealers can break even with both types and the market can clear.[23]

With this result in mind, we define $S_h \equiv \Delta V_h - p$ as the total surplus generated by a purchase, that is, the difference between the reservation value of a buyer, $\Delta V_h$, and the reservation value of a dealer, $p$. The intermediation fee is set so as to share the total surplus between the buyer, who receives $S_h - \phi_h$, and the dealer, who receives $\phi_h$. Symmetrically, let $S_\ell \equiv p - \Delta V_\ell$ denote the total surplus generated by a sale. The purchase and sale surpluses solve the following pair of Bellman equations:

$$rS_h = u_h - rp - \delta\pi_\ell (S_h + S_\ell) - \max_\theta \{\alpha(\theta)S_h - \gamma\theta\}, \tag{15}$$

$$rS_\ell = rp - u_\ell - \delta\pi_h (S_h + S_\ell) - \max_\theta \{\alpha(\theta)S_\ell - \gamma\theta\}. \tag{16}$$

To solve this system of equations for a given $p$, we define:

$$\Gamma(S) = (r + \delta)S + \max_\theta \{\alpha(\theta)S - \gamma\theta\}.$$

23. A related result occurs in the model of Gavazza (2011). There is a continuous distribution of private valuations of the asset across investors, and participation in the market is costly. As a result, there are two thresholds for private valuations: one above which agents participate as buyers and one below which they participate as sellers. Participation requires that agents get a positive surplus from a trade. So if trades were intermediated by dealers, as in DGP, the interdealer market price would satisfy the same condition as in our model.

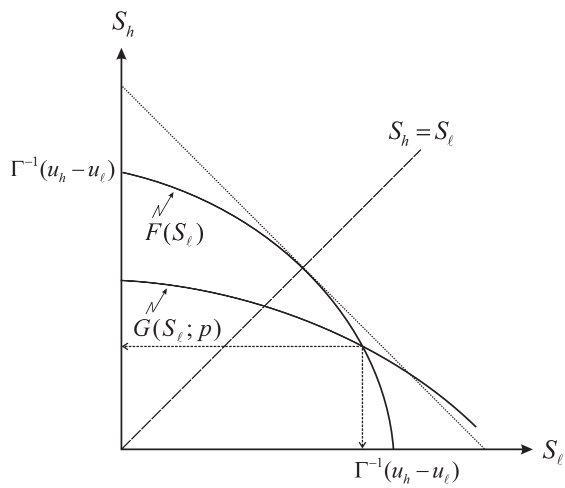


FIG. 1. The Determination of $S_\ell$ and $S_h$ Given $p$.

The function $\Gamma(S)$ is twice continuously differentiable, strictly increasing, strictly convex, and satisfies $\Gamma(0) = 0$ and $\Gamma(\infty) = 0$. Using this function, the system (15)–(16) can be written

$$\Gamma(S_h) + \Gamma(S_\ell) = u_h - u_\ell, \tag{17}$$

$$\Gamma(S_h) = u_h - rp + (1 - \delta\pi_\ell)S_h - \delta\pi_\ell S_\ell. \tag{18}$$

Equations (17) and (18) implicitly define two strictly decreasing, strictly concave, and continuously differentiable functions $S_h = F(S_\ell)$ and $S_h = G(S_\ell)$, respectively. The slopes of these functions are

$$F'(S_\ell) = -\frac{r + \delta + \alpha(\theta_\ell)}{r + \delta + \alpha(\theta_h)} < G'(S_\ell) = -\frac{\delta\pi_\ell}{r + \delta\pi_\ell + \alpha(\theta_h)},$$

where $\theta_\ell$ and $\theta_h$ are the maximizers of the programs defining $\Gamma(S_\ell)$ and $\Gamma(S_h)$, that is,

$$\alpha'(\theta_i)S_i = \gamma \quad \text{for } i \in \{\ell, h\}. \tag{19}$$

The ranking of these derivatives implies that if a solution to $F(S_\ell) = G(S_\ell)$ exists, then it must be unique. It also implies that $F(S_\ell)$ crosses $G(S_\ell)$ from above at any intersection, as illustrated in Figure 1. Using these results, we now address the issue of existence.

LEMMA 2. *There exist* $p_\ell < p_h$ *such that the system (15)–(16) has a strictly positive solution,* $[S_\ell(p), S_h(p)]$*, if and only if* $p \in (p_\ell, p_h)$*. This solution is unique. Moreover,*

*$S_\ell(p)$ is strictly increasing in $p$, $S_h(p)$ is strictly decreasing in $p$, and $S_\ell(p_\ell) = S_h(p_h) = 0$.*

According to Lemma 2, for a solution to exist the price must be neither too high nor too low, so that both buy orders and sell orders generate strictly positive surpluses. Moreover, the surplus from selling the asset is increasing in the price $p \in (p_\ell, p_h)$ while the surplus from buying the asset is decreasing in $p$. Recall, from (19), that there is less entry of dealers on a given side of the market if the corresponding trading surplus is lower. Thus, $\theta_h$ decreases with $p$ while $\theta_\ell$ increases with $p$.

*Market clearing.* We now solve for the distribution of investors across types and ensure that the asset market clears. First, since $p \in (\Delta V_\ell, \Delta V_h)$ from Lemma 1, the equilibrium buy- and sell-order flows must be equal:

$$\alpha(\theta_h)n_h(0) = \alpha(\theta_\ell)n_\ell(1). \tag{20}$$

The left-hand side is the buy-order flow originating from $(h, 0)$-type investors—investors with a high valuation for the asset but who do not own it—and the right-hand side is the sell-order flow originating from $(\ell, 1)$-type investors—investors who own the asset but have a low valuation for it. We derive the measure of buyers, $n_h(0)$, from the following steady-state condition:

$$\delta\pi_h n_\ell(0) = \delta\pi_\ell n_h(0) + \alpha(\theta_h)n_h(0). \tag{21}$$

The left-hand side is the inflow coming from $(\ell, 0)$ investors who change type. The first and second terms on the right-hand side are the outflows of type $(h, 0)$ investors who switch to low type with intensity $\delta\pi_\ell$ or who trade with intensity $\alpha(\theta_h)$, respectively. Similarly, the steady-state equation for the measure of $(\ell, 0)$-type investors is

$$\delta\pi_\ell n_h(0) + \alpha(\theta_\ell)n_\ell(1) = \delta\pi_h n_\ell(0),$$

which, from (20), can be rewritten as

$$\delta\pi_\ell n_h(0) + \alpha(\theta_h)n_h(0) = \delta\pi_h n_\ell(0). \tag{22}$$

Equations (21) and (22) can be interpreted as defining the ergodic distribution of a two-state Markov chain, for which $(h, 0)$-type investors transition to type $(\ell, 0)$ at Poisson rate $\delta\pi_\ell + \alpha(\theta_h)$, and $(\ell, 0)$-type investors transition to type $(h, 0)$ at Poisson rate $\delta\pi_h$. Together with the market-clearing condition, $n_h(0) + n_\ell(0) = 1 - A$, these steady-state conditions imply that

$$n_h(0) = (1 - A)\frac{\delta\pi_h}{\delta + \alpha(\theta_h)}. \tag{23}$$

By symmetry,

$$n_\ell(1) = A\frac{\delta\pi_\ell}{\delta + \alpha(\theta_\ell)}. \tag{24}$$

Thus, the equality of order flows, (20), can be rewritten as

$$(1-A)\frac{\delta\pi_h\alpha(\theta_h)}{\delta+\alpha(\theta_h)} = A\frac{\delta\pi_\ell\alpha(\theta_\ell)}{\delta+\alpha(\theta_\ell)}. \tag{25}$$

Recall from Lemma 2 that $\theta_h$ decreases with $p$ while $\theta_\ell$ increases with $p$. Therefore, one easily sees that the left-hand side of (25) is strictly decreasing and equal to zero at $p = p_h$, while the right-hand side is strictly increasing and equal to zero at $p = p_\ell$. Consequently, there exists a unique price $p \in (p_\ell, p_h)$ equating buy- and sell-order flows. Combining these observations with Lemma 1, we obtain the following result.

PROPOSITION 3. *There exists a unique competitive search equilibrium.*

The remaining equilibrium objects can be found using the same steps as in Section 2.2. In particular, the trading fees are

$$\phi_h = \frac{\theta_h\alpha'(\theta_h)}{\alpha(\theta_h)}S_h,$$
$$\phi_\ell = \frac{\theta_\ell\alpha'(\theta_\ell)}{\alpha(\theta_\ell)}S_\ell.$$

One recognizes here the Hosios (1990) condition for efficiency in economies with search frictions. The trading fee that a dealer charges for a buy order, $\phi_h$, is equal to the dealer's contribution to the sale surplus, as defined by the elasticity of the matching function. This elasticity determines the share of the trade surplus that a dealer appropriates, and hence it measures the implicit bargaining power of dealers when they trade. Note that unless the matching function has constant elasticity (such as in the Cobb–Douglas matching function) or unless $\theta_\ell = \theta_h$, the implicit bargaining power of dealers will differ depending on whether they trade with buyers or sellers. This asymmetry arises because intermediation on both sides of the market implies that there are two separate matching processes: one between dealers and buyers, and one between dealers and sellers. As a result, the Hosios condition must apply separately to each one of them.

*Comparative statics near the Walrasian limit.* In order to derive analytical comparative statics, it is useful to study the limiting equilibrium as search frictions vanish. To do this, we first assume that investors contact dealers at intensity $\lambda\alpha(\theta)$, and then we drive the search efficiency parameter, $\lambda$, to $+\infty$. We focus on the case $\pi_h > A$, so that high-valuation investors are on the long side of the market (a similar analysis applies to $\pi_h < A$). In this case, there are more high-valuation agents than assets. Hence, in the frictionless benchmark, the price adjusts so that high-valuation investors are indifferent between buying or not, and the price and allocations are

$$p^\star = \frac{u_h}{r}, \quad n_h^\star(0) = \pi_h - A, \quad \text{and} \quad n_\ell^\star(1) = 0.$$

The intermediation fees are implicitly equal to zero. Our main proposition characterizes equilibrium objects near this Walrasian limit.

PROPOSITION 4. *Let the matching technology be $\lambda\alpha(\theta)$ and assume that $\pi_h > A$. As $\lambda \to \infty$, the price, intermediation fees, and the measures of buyers and sellers admit the following first-order approximation:*

$$p = p^\star - \frac{\delta\pi_\ell s_\ell^\star}{\lambda r} + o\left(\frac{1}{\lambda}\right), \tag{26}$$

$$\phi_h = o\left(\frac{1}{\lambda}\right), \tag{27}$$

$$\phi_\ell = \frac{\theta_\ell^\star \alpha'(\theta_\ell^\star)}{\alpha(\theta_\ell^\star)} \frac{s_\ell^\star}{\lambda} + o\left(\frac{1}{\lambda}\right), \tag{28}$$

$$n_h(0) = n_h^\star(0) + \frac{A\delta\pi_\ell}{\lambda\alpha(\theta_\ell^\star)} + o\left(\frac{1}{\lambda}\right), \tag{29}$$

$$n_\ell(1) = n_\ell^\star(1) + \frac{A\delta\pi_\ell}{\lambda\alpha(\theta_\ell^\star)} + o\left(\frac{1}{\lambda}\right), \tag{30}$$

*where $o(1/\lambda)$ is such that* $\lim_{\lambda\to\infty} \lambda o(1/\lambda) = 0$, *while* $s_\ell^\star = \lim_{\lambda\to\infty} \lambda S_\ell$ *and* $\theta_\ell^\star = \lim_{\lambda\to\infty} \theta_\ell$ *jointly solve* $\max_\theta\{\alpha(\theta)s_\ell^\star - \gamma\theta\} = u_h - u_\ell$. *The market tightness for buyers satisfies* $\lim_{\lambda\to\infty} \theta_h = 0$, *and the purchase surplus satisfies* $\lim_{\lambda\to\infty} \lambda S_h = 0$.

As $\lambda \to \infty$, the different components of the competitive search equilibrium converge toward their Walrasian counterparts. Sellers, who are on the short side of the market, trade almost instantly: market tightness converges to some strictly positive limit, $\theta_\ell^\star$, so that the average execution time, $1/\lambda\alpha(\theta_\ell)$, converges to zero. Buyers, who are on the long side, trade with a nonzero asymptotic execution time. This sharp asymmetry in execution times is necessary to keep the buy- and sell-order flows balanced in the limit.

Proposition 4 allows for analytical comparative statics of various equilibrium objects of interest. For example, an increase in the gains from trade, $u_h - u_\ell$, causes the asymptotic sale surplus, $s_\ell^\star$, to increase. As a result, the asymptotic price discount, $\lim \lambda(p^\star - p)/p^\star$, increases because buyers expect that they will lose more utility upon switching to a low type. The increase in the sale surplus represents a profit opportunity for dealers. By free entry, the supply of intermediation services, $\theta_\ell^\star$, increases, which implies that investors can trade faster. As a result, the asymptotic measures of buyers and sellers, $n_h(0)$ and $n_\ell(1)$, decrease. By the same token, when $\theta_\ell^\star$ increases, intermediaries receive fewer trading opportunities on average, so the zero-profit condition requires that intermediation fees, $\phi_\ell^\star = \lim \lambda\phi_\ell$, go up.

An increase in the flow entry cost of dealers, $\gamma$, makes dealers more reluctant to post contracts and hence the asymptotic tightness, $\theta_\ell^\star$, decreases. As a result, the measures of buyers and sellers increase, the asymptotic sale surplus $s_\ell^\star$ increases, and so does the price discount. If the elasticity of the matching function is constant—for example, the Cobb–Douglas case—intermediation fees increase.

In response to an increase in the intensity of switching to the low valuation, $\delta\pi_\ell$, buyers anticipate that they may switch type soon after purchasing the asset. This lowers their willingness to pay for the asset, so the interdealer price decreases. Moreover, when investors switch to the low valuation more quickly after purchasing the asset, the asymptotic measure of sellers, $n_\ell(1)$, increases.[24]

Next, we consider the relationship between the price discount and the elasticity of the matching function. For this we focus on the special case of a Cobb–Douglas matching function.

COROLLARY 1. *Suppose $\alpha(\theta) = \theta^\eta$. Then,*

$$\theta_\ell^\star = \frac{\eta}{1-\eta}\frac{\Delta u}{\gamma} \quad \text{and} \quad s_\ell^\star = \left(\frac{\gamma}{\eta}\right)^\eta \left(\frac{u_h - u_\ell}{1-\eta}\right)^{1-\eta}.$$

*Thus, the asymptotic market tightness for sellers, $\theta_\ell^\star$, is an increasing function of $\eta$ and the asymptotic seller's surplus, $s_\ell^\star$, is a hump-shaped function of $\eta$. As a consequence, the seller's asymptotic trading fee and the asymptotic price discount are both hump-shaped functions of the matching elasticity, $\eta$.*

To gain intuition for the effect of the elasticity, $\eta$, note that the equation defining $s_\ell^\star$ can be written as

$$u_\ell + \max_\theta \{-\gamma\theta + \alpha(\theta)s\} = u_h = rp^\star.$$

This equation is an asymptotic indifference condition. It states that the asymptotic sale surplus, $s_\ell^\star$, must adjust so that a seller with a trading opportunity is indifferent between continuing search (on the left-hand side) and selling his asset (on the right-hand side).

One sees that, when $\eta \to 0$, then $\alpha(\theta) = \theta^\eta$ converges to 1 if $\theta > 0$ and to zero if $\theta = 0$. That is, an arbitrarily small market tightness, $\theta$, is sufficient to attain the maximum search intensity of 1. It is thus intuitive that, in equilibrium, market tightness is approximately zero and dealers incur almost no contract posting cost, $\alpha(\theta)$ is approximately 1, and the sale surplus is approximately equal to $u_h - u_\ell$.

When $\eta \to 1$, then the search technology becomes approximately linear. The seller's surplus must then converge to the search cost $\gamma$: if it were larger, the utility of continuing search would be unbounded, and if it were smaller it would be zero. Note that to keep the utility of continuing to search strictly positive, the market tightness must go to infinity.

24. One may have the impression that Proposition 4 implies the counterintuitive result that $n_h(0)$ increases with $\pi_\ell$. This, in fact, is not the case because $\pi_\ell$ also enters the leading term in the expansion: $n_h^\star(0) = \pi_h - A = 1 - \pi_\ell - A$. This leading term dictates that an increase in $\pi_\ell$ decreases $n_h(0)$, as intuition suggests.

We can now discuss comparative statics with respect to $\eta$. By the envelope condition, the derivative of the utility of continuing to search with respect to $\eta$ is equal to $s_\ell^\star \log(\theta_\ell^\star)(\theta_\ell^\star)^\eta$, which is negative for small $\theta_\ell^\star$ and positive for large $\theta_\ell^\star$. Graphically, an increase in elasticity "rotates" the search technology around the point (1, 1). It makes it less efficient for small $\theta$ and more efficient for large $\theta$.

Now assume that $\eta$ is small. Then in equilibrium $\theta$ is small, implying that an increase in $\eta$ will reduce search efficiency and, by the indifference condition, increase the sale surplus, $s_\ell^\star$. Vice versa, when $\eta$ is large, then in equilibrium $\theta$ is very large, implying that an increase in elasticity, $\eta$, will increase search efficiency and hence reduce the sale surplus, $s_\ell^\star$.

*Welfare.* Finally, we show that the unique competitive search equilibrium of our economy is also constrained efficient; that is, it maximizes society's welfare given the trading frictions arising from the matching technology. We define social welfare as the discounted sum of the utility flows of all investors net of dealers' entry costs,

$$\mathcal{W} = \int_0^{+\infty} e^{-rt} \left\{ n_{h,t}(1)u_h + n_{\ell,t}(1)u_\ell - \theta_{h,t}n_{h,t}(0)\gamma - \theta_{\ell,t}n_{\ell,t}(1)\gamma \right\} dt, \quad (31)$$

where we are using the fact that a social planner would not find it optimal to assign (costly) dealers to trade with $h$-valuation investors who hold assets (since an asset has no better use than in the hands of an $h$-type investor) or $\ell$-valuation investors who do not hold any asset (since an asset has no worse use than in the hands of an $\ell$-type investor). Intermediation fees do not enter $\mathcal{W}$ since they are pure transfers from investors to dealers that cancel out. The planner maximizes $\mathcal{W}$ with respect to time path for the state variables, $n_{i,t}(a)$, and controls, $\theta_{i,t}$, subject to the laws of motion for these state variables and to the balanced order flow condition,

$$\alpha(\theta_{h,t})n_{h,t}(0) = \alpha(\theta_{\ell,t})n_{\ell,t}(1). \quad (32)$$

If (32) were not satisfied, then the planner could reduce the measure of dealers on the long side of the market, thereby saving some entry costs. We omit the condition $\sum_{(i,a)} n_{i,t}(a) = 1$ as it is automatically satisfied by the laws of motion for $n_{i,t}(a)$ given (32) and the initial condition.

Proposition 5. *The competitive search equilibrium is constrained efficient.*

We establish this proposition in the Appendix, where we set up the Hamiltonian corresponding to this optimal control problem and check that its first-order conditions coincide with the conditions for a competitive search equilibrium. In particular, the costate variable associated with $n_{i,t}(a)$, which represents the marginal social value of an investor of type $i$ with holding $a$, coincides with $V_{i,t}(a)$, the private value of such an investor. The Lagrange multiplier associated with the balanced order-flow condition coincides with the equilibrium price, $p(t)$.

The proposition generalizes the welfare theorem for competitive search economies first derived by Moen (1997) to a two-sided OTC market augmented with a competitive interdealer market. In order to appreciate the significance of the previous proposition, consider the same economy with random search and *ex post* bargaining, as in DGP or LR. Let $\nu \in [0, 1]$ denote the dealer's bargaining power in the Nash bargaining to determine intermediation fees. Comparing the equilibrium conditions with the first-order conditions of the planner's problem (all shown in the Appendix) we obtain,

COROLLARY 2. *The equilibrium of an OTC market with random search and ex post bargaining is constraint efficient if and only if*

$$\frac{\alpha'(\theta_h)\theta_h}{\alpha(\theta_h)} = \frac{\alpha'(\theta_\ell)\theta_\ell}{\alpha(\theta_\ell)} = \nu. \tag{33}$$

Constrained efficiency in the opaque economy requires that dealers' bargaining power is equal to their contribution to the matching process as measured by the elasticity of the matching function. This is the generalization of the Hosios (1990) condition for efficiency in a two-sided random search market. Weill (2007) derived a related result in an economy without free entry but in which dealers can hold inventories. The corollary thus establishes that unless the matching function has a constant elasticity (as in the Cobb–Douglas matching function) or the two sides of the market have the same endogenous tightnesses, the Hosios condition (33) will not hold for a single bargaining power of dealers. So generically, the outcome of the opaque market will be socially inefficient.

*Transparency.* To close this subsection, consider an economy that is segmented into two distinct markets: a transparent market with a measure $\chi$ of investors who can observe contracts posted by dealers and direct their search accordingly, and an opaque market composed of a measure $1 - \chi$ of investors who search for dealers at random (as in Lester 2011). In contrast to investors who are assigned to a particular market, dealers are free to enter the transparent market or the opaque market. We call $\chi$ the degree of transparency of the overall economy. From the results above we have the following corollary.

COROLLARY 3. *The optimal degree of transparency is* $\chi = 1$.

Our economy highlights an important welfare-enhancing role of transparency: since investors essentially choose their trading speed, those investors with the largest gains from trade obtain fast execution, and those who have less to gain from trade ultimately execute their orders more slowly. This result provides support for recent regulatory changes (such as the TRACE system) that promote transparency in OTC markets.

### *3.2 Asset Holdings, Market Fragmentation, and Trading Costs*

In the previous section, we restricted investors' asset holdings in order to focus on their choice of trading speeds and the corresponding trading costs. In this section, we adopt a special matching technology, which implies that market tightness and average execution speeds are equal across active submarkets, and we focus on the equilibrium distribution of asset holdings and the relationship between trade size and trading costs.

More specifically, we adopt the same environment as the one described in Section 1 but we suppose that $\alpha(\theta) = \mu \min\{1, \theta\}$, with $\mu > 0$, so that there is a strict complementarity between the number of dealers and the number of orders processed. One interpretation of this technological assumption is that if $\theta > 1$, each order must be matched to an individual dealer, and the average time for a dealer to process an order is $1/\mu$. If $\theta < 1$, there are not enough dealers to handle all of the orders. In this case, orders are randomly assigned to dealers, so that each order is matched to a dealer with probability $\theta$.[25] For simplicity, we also assume that $\pi_{i,j} \equiv \pi_j$ for all $i, j \in \mathcal{I}^2$, with $\sum_{j \in \mathcal{I}} \pi_j = 1$, so that preference shocks are i.i.d. across investors.

*Active submarkets.* When an investor of type $i$ with asset holdings $a$ contacts a dealer and chooses a new portfolio $a'$, the utility gain is

$$\Lambda_i(a', a) = V_i^\star(a') - V_i^\star(a) - p(a' - a),$$

less any fee charged by the dealer. Therefore, as in the benchmark model, an investor of type $i$ with asset holdings $a$ chooses a submarket that solves

$$\max_{\theta, a', \phi} \left\{ \mu \min\{1, \theta\} \left[ \Lambda_i(a', a) - \phi \right] \right\}$$

subject to the free-entry condition of dealers

$$\mu \min\left\{1, \frac{1}{\theta}\right\} \phi \leq \gamma, \text{ with equality if } \theta > 0.$$

Substituting the constraint into the objective, the optimal choice of $\theta$ for a type $i$ investor with asset holdings $a$ is

$$\theta_i(a) = \begin{cases} 0 \\ [0, 1] \\ 1 \end{cases} \text{ if } \max_{a'} \Lambda_i(a', a) \begin{matrix} < \\ = \\ > \end{matrix} \gamma/\mu. \tag{34}$$

In words, given the prevailing asset price $p$, an investor of type $i$ with asset holdings $a$ will rebalance his portfolio in equilibrium if the (maximized) gain from doing so exceeds the expected cost incurred by a dealer to process the transaction, which is simply the flow cost $\gamma$ multiplied by the average time to execute the order $1/\mu$. Note that dealers' entry generates no congestion on other dealers—their matching rate is

25. It should be noted that this matching technology does not satisfy the assumptions imposed earlier: it is not strictly increasing and continuously differentiable everywhere, and $\alpha'(0) = \mu < \infty$. In Appendix A.11, we verify that the properties of equilibrium reported in Section 2 remain true under this specification.

$\mu$—as long as $\theta \leq 1$. As a result, active submarkets will typically have $\theta_i(a) = 1$. In each of these active submarkets, a contract is posted with a fee,

$$\phi_i = \phi \equiv \frac{\gamma}{\mu}, \tag{35}$$

which ensures that dealers earn zero profits, along with a quantity $q_i(a) = a_i' - a$, where

$$a_i = \arg\max_{a' \geq 0} \ \Lambda_i(a', a). \tag{36}$$

Taken together, the results above imply that the Bellman equation for an investor of type $i$ with asset holdings $a$ can be rewritten as

$$rV_i^\star(a) = u_i(a) + \delta \sum_k \pi_k \left[V_k^\star(a) - V_i^\star(a)\right] + \mu \max\left\{0, \max_{a'}\left[\Lambda_i(a', a) - \frac{\gamma}{\mu}\right]\right\}.$$

From the viewpoint of an investor, it is as if trading opportunities arrive at rate $\mu$, and they can choose whether or not to trade at a cost $\phi = \gamma/\mu$. In this sense, our model is formally equivalent to a model of trade with fixed adjustment costs, as in Lo, Mamaysky, and Wang (2004), though the costs of rebalancing one's portfolio arise endogenously in our model.

As in most models with adjustment costs, the equilibrium in our environment could involve "inaction regions," where an investor of type $i$ who holds $a_j$ for $j \neq i$ chooses not to submit a trade order because the fee required to compensate a dealer for executing this order exceeds the expected gains to the investor. In what follows, we will focus on equilibria in which investors who hold their optimal portfolio *always* rebalance their asset holdings after receiving a preference shock. As we establish later, this allows us to sidestep the calculation of value functions for all $a \in \mathbb{R}_+$, and instead characterize equilibrium asset holdings using a simple variational argument.

*Optimal asset holdings.* Before characterizing the asset demand functions in our candidate equilibrium, we sketch the logic of our approach (formal proofs are in Appendix A.11). Let $N_i \subset \mathbb{R}_+$ denote the *inaction region* for an investor of type $i$, that is, the set of values of $a$ such that an investor of type $i$ with asset holdings in this set would choose not to trade. Given $\gamma/\mu$ sufficiently small, we can establish that $\bigcap_{i \in \mathcal{I}} N_i = \emptyset$, so that an investor of type $i$ with $a \in N_i$ will surely trade if he receives a preference shock $j \neq i$. Given this behavior, there exists a unique $a_i \in N_i$ that maximizes the payoffs of a type $i$ investor.

To characterize these optimal asset holdings, note that $a_i$ is a solution to (36), for each $i \in \mathcal{I}$, where

$$rV_i^\star(a) = u_i(a) + \delta \sum_k \pi_k \left[V_k^\star(a) - V_i^\star(a)\right] \tag{37}$$

if $a \in N_i$, and

$$rV_i^\star(a) = u_i(a) + \delta \sum_k \pi_k \left[V_k^\star(a) - V_i^\star(a)\right] + \mu \max_{a'} \left[\Lambda_i(a', a) - \frac{\gamma}{\mu}\right] \quad (38)$$

otherwise. One can manipulate (37) and (38) to see that, for all $a$ in the inaction region $N_i$, the net benefit to a type $i$ investor from acquiring portfolio $a$, $V_i(a) - pa$, is a positive affine transformation of

$$\frac{(r+\mu)\left[u_i(a) - rpa_i\right] + \delta \sum \pi_k \left[u_k(a) - rpa\right]}{(r+\mu)(r+\delta) - \delta\mu\pi_j}. \quad (39)$$

Since any optimal asset holding must belong to the inaction region $N_i$, it follows that $a_i$ must satisfy the first-order condition for $V_i(a) - pa$ to be maximized:

$$\frac{(r+\mu)\, u_i'(a_i) + \delta \sum \pi_k u_k'(a_i)}{r+\mu+\delta} = rp. \quad (40)$$

Note that this equation uniquely determines the optimal asset holding for an investor of type $i$, $a_i$. Studying the left-hand side of (40), we see that the demand of an investor of type $i$ depends on his current marginal utility and his expected future marginal utility, weighted by the parameters $r$, $\mu$, and $\delta$. In particular, an investor places more weight on his current value of marginal utility when he is impatient, when orders are executed quickly, and when he expects to stay in his current state for a long time.[26]

Moreover, if $u_i'(a_i) > \sum_k \pi_k u_k'(a_i)$, then $\partial a_i/\partial\mu > 0$; that is, the investor's demand for the asset increases as trading frictions are reduced. Intuitively, when an investor's current marginal utility is relatively high, an increase in trading speed will cause him to take a larger position now since it will be easier for him to unwind this position in the event of an adverse preference shock. Conversely, if $u_i'(a_i) < \sum_k \pi_k u_k'(a_i)$, that is, if the investor's marginal utility in the current state is below average, then $\partial a_i/\partial\mu < 0$. More generally, as in LR, faster trading causes investors to take more extreme positions, while slower trading gives agents incentives to choose a more moderate portfolio that will not be too far from the optimal after any preference shock. Finally, as $\mu$ goes to infinity, the optimal portfolio tends to the value of $a_i$ that solves $u_i'(a_i) = rp$, which is the portfolio choice that would prevail in a competitive market where all trades can be executed instantaneously.

26. The asset demand functions characterized in (40) are closely related to the asset demand functions derived in LR, where investors must bargain with dealers over the quantity to be traded and the corresponding intermediation fee. In particular, our asset demand functions correspond to those of LR *when dealers have zero bargaining power*. Interestingly, LR show that investors' portfolio decisions are socially inefficient whenever dealers have positive bargaining power. Hence, the competitive search mechanism studied in this paper corrects the inefficiencies that arise from an environment with random search and bargaining; intuitively, competition between dealers for order flow ensures that investors receive exactly their marginal gain from readjusting their portfolio.

*The steady-state distribution and market-clearing conditions.* Given the results mentioned above, asset holdings in the candidate equilibrium can be described by a list $\{a_i\}_{i=1}^I$. Letting $n_{ij}$ denote the measure of agents of type $i$ with asset holdings $a_j$, the steady-state distribution must satisfy

$$\delta\pi_i \sum_{k\in\mathcal{I}} n_{kj} - \delta n_{ij} - \mu n_{ij} = 0, \text{ for } i \neq j, \tag{41}$$

$$\delta\pi_i \sum_{k\in\mathcal{I}} n_{ki} + \mu \sum_{k\neq i} n_{ik} - \delta n_{ii} = 0, \text{ for } i = j, \tag{42}$$

$$\sum_{(i,j)\in\mathcal{I}^2} n_{ij} = 1. \tag{43}$$

The first term in equation (41) represents the inflow of agents into state $ij$, with $i \neq j$, which only occurs because of type switching. The second two terms in equation (41) represent the outflow from state $ij$, which occurs because of both type switching (the second term) and trade (the third term). Equation (42) is the steady state condition for investors holding the optimal portfolio. In this case, inflow can occur either when investors successfully rebalance their portfolio or when they (luckily) receive a preference shock that corresponds to their current asset holdings. Outflow, of course, occurs because of type switching.

Solving, we find that $\sum_j n_{ij} = \sum_j n_{ji} = \pi_i$, so that

$$n_{ij} = \frac{\delta\pi_i\pi_j}{\mu+\delta}, \quad \text{for } j \neq i, \tag{44}$$

$$n_{ii} = \frac{\delta\pi_i^2 + \mu\pi_i}{\mu+\delta} \quad \text{for } j \neq i. \tag{45}$$

Note that the distribution of probabilities across states is symmetric, $n_{ij} = n_{ji}$. Also, $\partial n_{ij}/\partial\mu < 0$ and $\partial n_{ij}/\partial\delta > 0$ if $j \neq i$, while $\partial n_{ii}/\partial\mu > 0$ and $\partial n_{ii}/\partial\delta < 0$: the measure of investors who are matched to their desired portfolio increases with the speed of execution and decreases with the arrival rate of preference shocks.

Turning now to the determination of the asset price in the competitive interdealer market, we note that market clearing requires $\sum_{i,j} n_{ij}a_i = A$. Using the fact that $\sum_j n_{ij} = \pi_i$, this condition reduces to

$$\sum_i \pi_i a_i = A. \tag{46}$$

*Equilibrium and comparative statics.* An equilibrium such that $\theta_i(a_j) = 1$ for all $i \neq j$ can thus be defined as a list $\{a_i\}$, $\{n_{ij}\}$, $\phi$, and $p$ that satisfy (35), (40), and (44)–(46). The individual portfolio choices $(a_j)$ in (40) depend on $p$, the equilibrium price in the interdealer market. The distribution of investors over portfolios and preference types is given by (44) and (45). Given these individual demands, the market-clearing condition (46) determines a unique price. Finally, the choice of $\theta_i(a_j)$ described in (35) is optimal provided that $\gamma/\mu$ is sufficiently small.

Using these equilibrium conditions, we can explore the model's implications for prices, trading activity, and allocations. For example, although we have disentangled the asset price from the intermediation fee, one can still compute the *effective price* that an investor pays (or receives) per unit of the asset he buys (or sells). 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)

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

per unit of the asset. The difference between the prices at which investors buy and sell is sometimes treated as a measure of market liquidity. Notice that if $a_j - a_i > 0$, then

$$\hat{p}_{ij} - \hat{p}_{ji} = \frac{2\phi}{a_j - a_i} = \frac{2\gamma}{\mu\left(a_j - a_i\right)} > 0. \tag{47}$$

So for this typical "round-trip" transaction, investors trade at a higher effective price when they buy than when they sell. One can notice from (47) that the effective spread decreases with the size of the order, which is in accordance with the evidence documented by Schultz (2001) and Edwards, Harris, and Piwowar (2007) for OTC markets.[27]

A decrease in $\gamma$, the operating costs for dealers, also causes a decrease in the bid–ask spread. The effects of a change in $\mu$ are less obvious. On the one hand, an increase in $\mu$ has a direct effect in reducing $\phi = \gamma/\mu$. On the other hand, a change in $\mu$ also affects equilibrium asset holdings and hence the quantity traded, $a_j - a_i$. However, since the distribution of asset holdings spreads out as $\mu$ increases, the quantity of assets traded in many individual trades tends to increase (see LR); in the case of $I = 2$, for example, it can be checked that $|a_2 - a_1|$ increases with $\mu$, so the bid–ask spread unambiguously decreases with $\mu$.

The model also has implications for trading volume, which is defined as follows. The flow of investors who can readjust their portfolios per unit of time is $\mu$. A fraction

27. Similarly, Green, Hollifield, and Schürhoff (2007) document that dealers earn lower average markups on larger trades in the market for municipal bonds. By estimating a bargaining model, they find that dealer's bargaining power is substantial and it decreases in trade size.

$n_{ij}$ of these investors readjust their portfolio from $a_i$ to $a_j$ so that the quantity they trade is $|a_j - a_i|$. Thus, the total *volume of trade* is

$$\mathcal{V} = \frac{\mu}{2} \sum_{i,j} n_{ij} |a_j - a_i|. \tag{48}$$

An increase in $\mu$ has three distinct effects on trade volume. First, the measure of investors in any individual state $(i, j) \in I^2$ who can readjust their portfolios 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, as discussed, the distribution of asset holdings spreads out, which tends to increase the quantity of assets traded in many individual trades, and hence increase volume. Given (44) and (48), it is easy to check that the first two effects combined lead to an increase in $\mathcal{V}$. Therefore, when the third effect is also positive the total volume of trade unambiguously increases with $\mu$.

We can also study some limiting cases. First, as $\mu \to \infty$, so that trades are executed instantaneously, we see from (35) that intermediation fees go to 0. Moreover, from (40), the individual demand for the asset converges to the Walrasian demand, $u'_j(a_j) = rp$, and from (44) and (45) all investors hold their desired portfolios. The same is true as $\delta \to 0$, since type never changes. Finally, as $\gamma \to 0$, the intermediation fee $\phi \to 0$. However, a decrease in $\gamma$ will only have an effect on asset prices and allocations if it causes a change in the set of active submarkets; if the initial value of $\gamma$ was sufficiently small, so that $\theta_i(a_j) = 1$ for all $i \neq j$, then a decrease in $\gamma$ has no effect on the allocation of assets across agents or on asset prices.

## 4. A NUMERICAL EXAMPLE

In this final section, we return to the general environment described in Section 1 and develop a numerical example to show that our model generates quantitative outcomes of reasonable magnitude, in terms of trade size, execution speed, and trading costs. We then use this example to offer some numerical comparative statics. We assume that preference shocks are multiplicative with an isoelastic utility function, $u_i(a) = \varepsilon_i v(a)$ with $v(a) = a^{1-\sigma}/(1-\sigma)$, and the matching function is Cobb–Douglas, $\alpha(\theta) = \lambda\theta^{\eta}$. The baseline values, shown in Table 1, are chosen as follows.

First, as is common, we set the discount rate to 5% per year. Second, given that our preference specification is homogenous, we normalize the asset supply to $A = 1$.[28] Third, with Cobb–Douglas matching, since the unit used to measure market tightness

28. Up to an appropriate scaling of the contract posting cost, the only effect of scaling $A$ up and down is to scale down all relevant equilibrium objects by the same constant. Namely, if $\{p, a_i, \theta_i\}$ is an equilibrium when the asset supply is $A$ and the contract posting cost is $\gamma$, then, for any $\lambda > 0$, $\{\lambda^{-\sigma} p, \lambda a_i, \theta_i\}$ is an equilibrium when the asset supply is $\lambda A$ and the contract posting post is $\gamma\lambda^{1-\sigma}$.

TABLE 1

Parameter Values

| Parameter | | Value |
|---|---|---|
| Asset supply | $A$ | 1 |
| Cost of posting contract | $\gamma$ | 1 |
| Contact intensity | $\lambda$ | 85 per year |
| Elasticity of matching function | $\eta$ | 0.7 |
| Discount rate | $r$ | 0.05 per year |
| Type-switching intensity | $\delta$ | 0.7 pear year |
| Elasticity of utility function | $\sigma$ | 4.75 |
| Number of utility types | $I$ | 30 |
| Transition probabilities | $\pi_i$ | $1/I$ |
| Utility types | $\varepsilon_i$ | $e^{-10+12.3\Phi^{-1}[0.05+0.90(i-1)/I]}$ |

Notes: In the last line, $\Phi^{-1}(q)$ is the inverse cumulative function of a standard normal distribution. According to the formula, the grid of $\{\omega_i\}_{i\in\{1,\ldots,I\}}$ is made up of 30 percentiles, equally spaced between 0.05 and 0.95, of a log normal distribution with mean $\mu = -10$ and standard deviation $\sigma = 12.3$.

is meaningless in our model, we can normalize the contract posting cost to 1 (see Shimer 2005 for a similar argument in his calibration of the Mortensen–Pissarides model).

The contact intensity, $\lambda$, is chosen to generate reasonable trading intensities, measured as follows. In a stationary equilibrium, the aggregate quantity of assets waiting to be either purchased or sold at any point in time is equal to $\sum_{ij} n_i(a_j)|a_i - a_j|$. The probability that a randomly chosen unit waiting to be traded belongs to the order of agent $(i, j)$ is

$$\psi_i(a_j) = \frac{n_i(a_j)|a_i - a_j|}{\sum_{k,\ell} n_k(a_\ell)|a_k - a_\ell|}.$$

Thus, the trading intensity of a randomly chosen unit is $\bar{\lambda} = \sum_{i,j} \psi_i(a_j)\lambda\theta_i(a_j)^\eta$. In our numerical example, the contact intensity, $\lambda = 85$, implies that the average trading intensity is $\bar{\lambda} = 286$; that is, it takes on average 0.87 days to trade one unit of asset.

We pick the rest of the parameters based on empirical targets from the municipal bond market. In our numerical example, the type switching intensity, $\delta = 0.7$, generates a turnover of about 46% per year, which is in the ballpark of the 56% turnover measured by Green, Hollifield, and Schürhoff (2007) for seasoned bonds over the 2000–04 time period. We adopt a parsimonious parameterization for the distribution of utility types in order to obtain a rough match of the distribution of relative trade sizes, as proxied by the distribution of relative sale sizes from customers to dealers reported by the Fact Book of the Municipal Securities Rulemaking Board (2008). Figure 2 shows the theoretical distribution generated by the model and compares it to its empirical counterpart.

Finally, to set the elasticity of the matching function, $\eta$, and the elasticity of the utility function, $\sigma$, we consider two empirical targets on transaction costs. The first empirical target is the average transaction cost per unit traded, which was between 50 bp and 100 bp during the 2000–04 period according to the markup data reported

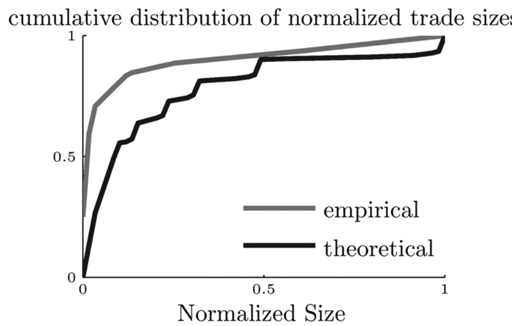


FIG. 2. Empirical versus Theoretical CDFs of (Normalized) Trade Sizes.

NOTES: The top line plots the empirical distribution of normalized trade sizes reported by the Fact Book of the Municipal Securities Rulemaking Board (2008). The bottom line is the corresponding distribution of relative trade sizes generated by our calibrated model.

by Green, Hollifield, and Schürhoff (2007). The average unit transaction cost that an econometrician would measure in a stationary equilibrium of our model is

$$\sum_{ij} \rho_i(a_j) \frac{\phi_i(a_j)}{|a_j - a_i|},$$

where

$$\rho_i(a_j) = \frac{n_i(a_j)\lambda\theta_i(a_j)^{1-\eta}}{\sum_{k,\ell} n_k(a_\ell)\lambda\theta_k(a_\ell)^{1-\eta}}$$

is the probability that a randomly chosen trade originates from an investor of utility type $i$ with asset holdings $a_j$. In our numerical example, the average unit transaction cost is 38 bp, smaller but comparable to the markups measured in the data. The other empirical target we consider is the relationship between transaction cost and trade size. According to the regressions of Harris and Piwowar (2004), multiplying the trade size by 10 divides the unit transaction cost by 2, corresponding to an elasticity of about $-\log(2)/\log(10) = -0.3$. In a stationary equilibrium of our model, this elasticity can be measured by calculating the theoretical coefficient of a univariate regression of log unit transaction cost on log trade size:

$$\frac{\text{cov}\left[\log\left(\frac{\phi_i(a_j)}{|a_j - a_i|}\right), \log(|a_j - a_i|)\right]}{\text{var}\left[\log(|a_j - a_i|)\right]},$$

where the covariances and variances are calculated based on the probabilities $\rho_i(a_j)$. In our numerical example, this elasticity turns out to be $-0.29$.

We conclude this section with comparative statics of transaction costs with respect to two parameters: the trading intensity, $\lambda$, and the elasticity of the matching function, $\eta$. From the left-hand side, the first plot in Figure 3 shows that increasing the contact intensity, $\lambda$, reduces the equilibrium level of transaction costs, which is consistent

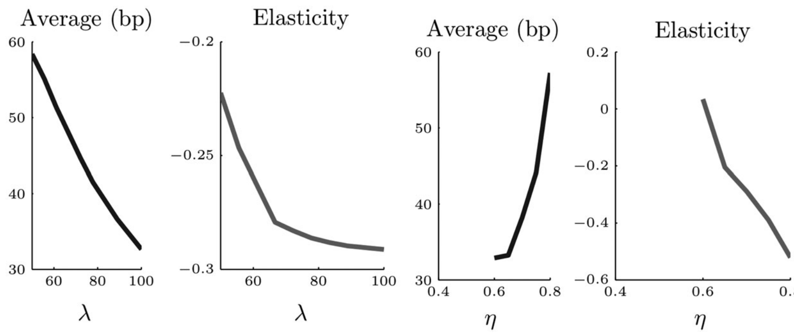


FIG. 3. Impact of λ and $\eta$ on (i) Average Unit Fee and (ii) Elasticity of Unit Fee to Trade Size.

with the theoretical result that we derived earlier in the model with fixed trade size. The second panel reveals that trading intensity also has an impact on the elasticity of transaction costs to trade size: as λ increases, transaction costs respond more strongly to changes in trade size. The third panel of the figure shows that increasing the elasticity of the matching function, $\eta$, tends to increase unit trading fees. This is in line with the observation that in a competitive search equilibrium, $\eta$ represents the share of the surplus that is appropriated by dealers. It is natural, then, that a larger $\eta$ is associated with larger transaction costs. The fourth and last panel shows that there is a sharp negative relationship between the elasticity of the matching function, $\eta$, and the elasticity of transaction costs to trade size. Together with the theoretical analysis above, this finding suggests that the empirical relationship between unit fee and trade size can be useful to infer properties of the matching function.

## 5. CONCLUSION

We have developed a model of a two-sided asset market where trades are intermediated by dealers and involve a time-consuming matching process. In contrast to the description of OTC markets in much of the existing literature, we assume that dealers post terms of trade in order to attract order flow, and investors observe these terms of trade and direct their orders to the dealer who promises the maximum expected payoff. This formulation is a good description of many OTC markets, where prices, trading costs (e.g., bid–ask spreads), and execution times are either readily observable or they are made public shortly after trades occur, which enables dealers to achieve commitment through reputation.

We characterize the equilibrium in a general setting, and illustrate how the model can account for several important trading patterns in OTC markets, which do not emerge from existing models. We then study two special cases that allow us to highlight the differences between our model and the existing literature. Finally, we calibrate our model, illustrate that it generates reasonable quantitative outcomes, and use it to study how trading frictions affect the per unit trading costs that investors pay in equilibrium.

## APPENDIX: PROOFS

### *A.1 Proof of Proposition 1*

We start by noting that $V_i(a, \sigma, \theta)$ can be written as

$$V_i(a, \sigma, \theta) = V_i(a, \mathbf{0}, 0) + \frac{\alpha(\theta)}{r + \delta + \alpha(\theta)} \left[ V_i^\star(a + q) - pq - \phi - V_i(a, \mathbf{0}, 0) \right], \quad \text{(A1)}$$

where

$$V_i(a, \mathbf{0}, 0) = \frac{u_i(a) + \delta \sum_{j=1}^{I} \pi_{ij} V_j^\star(a)}{r + \delta}.$$

Equation (A1) shows that the value of submitting an order for contract $\sigma$ when the market tightness is $\theta$ is the sum of two terms: the first term is the no-trade utility, $V_i(a, \mathbf{0}, 0)$, and the second term is proportional to the trading surplus, $V_i(a + q) - pq - \phi - V_i(a, \mathbf{0}, 0)$. The constant of proportionality, $\frac{\alpha(\theta)}{r+\delta+\alpha(\theta)} \leq 1$, captures the time discounting of trading delays associated with tightness $\theta$. The next result is then immediate.

LEMMA A1. *In any equilibrium, $V_i^\star(a) \geq V_i(a, \mathbf{0}, 0)$. Moreover the function $\theta \mapsto V_i(a, \sigma, \theta)$ has the following properties:*

*(i) If $V_i^\star(a + q) - pq - \phi \leq V_i(a, \mathbf{0}, 0)$, then $V_i(a, \sigma, \theta)$ is decreasing in $\theta$.*
*(ii) If $V_i^\star(a + q) - pq - \phi > V_i(a, \mathbf{0}, 0)$, then $V_i(a, \sigma, \theta)$ is strictly increasing in $\theta$.*

The first part of the lemma is clear: since investors can always choose $\mathbf{0}$, their maximum attainable utility must be at least equal to $V_i(a, \mathbf{0}, 0)$. The second part of the lemma simply asserts that if there are strict gains from trading $\sigma$, $V_i^\star(a + q) - pq - \phi > V_i(a, \mathbf{0}, 0)$, then investors have a strict preference for higher tightness, because it leads to smaller trading delays. Otherwise, they prefer not to send orders for $\sigma$.

Now turning to the proof of Proposition 1, we first show that the right-hand side of (7) is an upper bound for $V_i^\star(a)$. Indeed, for $\sigma = \mathbf{0}$, $V_i(a, \mathbf{0}, 0)$ is equal to the right-hand side of (7) evaluated at $\theta = 0$. For $\sigma \in \Sigma^\star$ but different from $\mathbf{0}$, the dealer's zero-profit condition writes $\alpha[\Theta(\sigma)]\phi = \gamma\Theta(\sigma)$. Thus,

$$V_i(a, \sigma, \Theta(\sigma)) = \frac{u_i(a) + \delta \sum_{j \in \mathcal{I}} \pi_{ij} V_j^\star(a) + \alpha[\Theta(\sigma)] \left[ V_i^\star(a + q) - pq \right] - \gamma\Theta(\sigma)}{r + \delta + \alpha[\Theta(\sigma)]},$$

which is clearly less than the right-hand side of (7). Thus, taking the sup over all $\sigma \in \Sigma^\star$, we find $V_i^\star(a)$ is less than the right-hand side of (7).

Toward a contradiction, suppose that $V_i^\star(a)$ is strictly less than the right-hand side of (7). Then, there must be some $(\hat{\theta}, \hat{q})$ such that

$$V_i^\star(a) < \frac{u_i(a) + \delta \sum_{j \in \mathcal{I}} \pi_{ij} V_j^\star(a) + \alpha(\hat{\theta}) \left[V_i^\star(a + \hat{q}) - p\hat{q}\right] - \gamma\hat{\theta}}{r + \delta + \alpha(\hat{\theta})}$$

$$\iff \quad V_i^\star(a) < \frac{u_i(a) + \delta \sum_{j \in \mathcal{I}} \pi_{ij} V_j^\star(a) + \alpha(\hat{\theta}) \left[V_i^\star(a + \hat{q}) - p\hat{q} - V_i^\star(a)\right] - \gamma\hat{\theta}}{r + \delta}. \tag{A2}$$

Note, using the lower bound $V_i(a, \sigma, \theta) \geq V_i(a, \mathbf{0}, 0)$ in inequality (A2), that $\hat{\theta} > 0$. Consider, then, the contract $\hat{\sigma} = (\hat{q}, \hat{\phi})$, where $\hat{\phi} = \frac{\gamma\hat{\theta}}{\alpha(\hat{\theta})} + \varepsilon$, for some $\varepsilon$ small enough so that $V_i(a, \hat{\sigma}, \hat{\theta}) > V_i^\star(a)$. Using again the lower bound in inequality (A2), one sees that $V_i^\star(a + \hat{q}) - p\hat{q} - \hat{\phi} > V_i^\star(a) \geq V_i(a, \mathbf{0}, 0)$. Thus, by Lemma A1, $V_i(a, \hat{\sigma}, \theta)$ is strictly increasing in $\theta$. Since it is also continuous, it then follows that $\Theta(\hat{\sigma}) < \hat{\theta}$. Given that $\alpha(\theta)/\theta$ is decreasing, we find that $\Pi(\hat{\sigma}) > 0$, which contradicts dealers' zero-profit condition.

Finally, for the second part of the proposition, we apply the contraction mapping to the auxiliary Bellman equation (7). Consider some arbitrary lower and upper bounds for the price, $0 < \underline{p} < \overline{p}$. If asset holdings are unrestricted, let $\overline{a}$ and $\underline{a}$ be the solutions to $u_I'(a) = r\underline{p}$ and $u_1'(a) = r\overline{p}$, respectively. Since $u_i'(a) \leq rp$ for all $a \geq \overline{a}$ and all $i$, and $u_i'(a)(\overline{p}) \geq rp$ for all $a \leq \underline{a}$ and all $i$ and all $p \in [\underline{p}, \overline{p}]$, one can show that an investor will always find it optimal to choose holdings $a \in [\underline{a}, \overline{a}]$.

If asset holdings are restricted, we let $\underline{a} = \min \mathcal{A}$ and $\overline{a} = \max \mathcal{A}$: by assumption, asset holdings always lie in the interval $[\underline{a}, \overline{a}]$. To derive bounds on the choice of $\theta$, note that up to a positive constant of proportionality, the left derivative with respect to $\theta$ of the equation to be maximized on the right-hand side of (7) is equal to

$$-(r+\delta)\gamma + \left[\alpha'(\theta^-)\theta - \alpha(\theta)\right]\gamma$$
$$+\alpha'(\theta^-)(r+\delta)\left[V_i^\star(a+q) - pq - \frac{u_i(a) + \sum \pi_{ij} V_j^\star(a)}{r+\delta}\right].$$

The first term is strictly negative and independent of $\theta$. The second term is negative because $\alpha(\theta)$ is concave. In the third term, the square bracket is bounded independently of $\theta$ and $p \in [\underline{p}, \overline{p}]$.[29] Given that $\alpha'(\infty) = 0$, this implies that there

29. Indeed, $V_i(a) \leq \max_{i \in \mathcal{I}} \frac{u_i(\overline{a})}{r} + \overline{pa}$, so that the maximum attainable utility is less than the present value of the maximum utility flow from holding the maximum quantity of assets, plus the time-zero value of selling the maximum quantity of assets at the maximum price. To see this, note that the intertemporal value of an investor has two terms. The first term is the expected present value of utility flows, $u_i[a(t)]$, net of search cost, $-\gamma\theta(t)$, which is clearly less than $\frac{u_I(\overline{a})}{r}$. The second term is the expected present value of the benefits of selling minus the costs of buying assets at the random contact times $0 < T_1 < T_2 < \ldots$. For given

exists some $\overline{\theta}$ such that, for all $\theta > \overline{\theta}$, all $p \in [\underline{p}, \overline{p}]$, and all $(i, a) \in \mathcal{I} \times [\underline{a}, \overline{a}]$, the left-derivative of equation (7) is strictly negative. Therefore, an investor will always find it optimal to choose a market tightness $\theta \in [0, \overline{\theta}]$.

Let $S = \mathcal{I} \times [\underline{a}, \bar{a}] \times [\underline{p}, \overline{p}]$ when assets are unrestricted, and let $S = \mathcal{I} \times \mathcal{A} \times [\underline{p}, \overline{p}]$ when assets are restricted. Let $\mathcal{C}(S)$ be the space of bounded, continuous functions $f : S \to \mathbb{R}$ equipped with the sup norm. Notice that $\mathcal{C}(S)$ is a complete normed vector space. The right-hand side of (7) defines an operator $T$:

$$T(V_i)(a, p) = \max_{(a',\theta)\in[\underline{a},\overline{a}]\times[0,\overline{\theta}]} \left[ \frac{u_i(a) + \delta \sum_{j\in\mathcal{I}} \pi_{ij} V_j(a) + \alpha(\theta)[V_i(a') - p(a' - a)] - \gamma\theta}{r + \delta + \alpha(\theta)} \right]. \quad \text{(A3)}$$

If $V \in \mathcal{C}(S)$, then, from the Theorem of the Maximum $T(V) \in \mathcal{C}(S)$. See theorem 3.6 in Stokey and Lucas (1989). Furthermore, $T$ is monotonic and $T(f + k)(a, i) \leq T(f)(a, i) + \frac{\delta+\alpha(\overline{\theta})}{r+\delta+\alpha(\overline{\theta})} k$ for all $k \geq 0$ and all $f \in \mathcal{C}(S)$. Therefore, from Blackwell's Theorem, $T$ is a contraction on $\mathcal{C}(S)$ and it has a unique fixed point in $\mathcal{C}(S)$. See theorem 3.3 in Stokey and Lucas (1989). Finally, if $V_i(a, p)$ is a strictly increasing function, then $T(V_i)(a, p)$ is also strictly increasing so that the unique fixed point of $T$ is increasing. But since $u_i(a)$ is strictly increasing, the Bellman equation (7) implies that $V_i(a)$ is the sum of a strictly increasing function and an increasing function, and so is strictly increasing as well. Finally, since we considered arbitrary lower and upper bounds, $\underline{p}$ and $\overline{p}$, and since $\overline{a} \to \infty$ as $\underline{p} \to 0$, and $\underline{a} \to 0$ as $\overline{p} \to \infty$, the properties extend by continuity to the domain $S = \mathcal{I} \times \mathcal{A} \times (0, \infty)$. □

*A.2 Proof of Proposition 2*

We first note that, by Theorem of the Maximum, the correspondence $(a_i(a, p), \theta_i(a, p))$ maximizing (A3) is compact valued and upper hemi continuous. Moreover, one sees that $a_i(a, p)$ is independent on $a$: it is equal to arg $\max_{a'\in[\underline{a},\overline{a}]} V_i(a', p) - pa'$ if $0 \notin \theta_i(a, p)$ and to $[\underline{a}, \overline{a}]$ otherwise. With this in mind, let

$$\mathcal{M}(p) \equiv \left\{(a_i(p), \theta_j(a_i(p), p)), (i, j) \in \mathcal{I}^2\right\}.$$

The correspondence $\mathcal{M}(p)$ contains the collection of candidate equilibrium asset holdings if the price is $p$. It is clearly compact valued and is easily shown to be upper hemi continuous. As explained in the text, we look for an equilibrium in which the distribution of types has a finite support. To deal with potential nonconvexity arising if the Bellman equation has more than one maximizer, we partition the population of investors into $G \geq 2$ groups of size $\mu^g$, for some $\mu^g$ in the $G$-dimensional simplex, $\Delta^G$. We assume that each investor in group $g$ optimally keeps its asset holdings in the

realization of contact times, this present value can be written as $a(0)pe^{-rT_1} + \sum_{n=1}^{\infty} pa(T_n)[e^{-rT_{n+1}} - e^{-rT_n}]$. Indeed, the investor resells $a(T_n)$ at time $T_{n+1}$ after buying it at time $T_n$. Clearly, the first term is less than $\overline{p}\bar{a}$, and all terms in the infinite sum are negative.

support $\{a_1^g, a_2^g, \ldots, a_I^g\}$, where for all $i$ and $g$, $a_i^g$ is maximizing (7). If the investor has utility type $i$ but asset holding $a_j^g$, $j \neq i$, then he sends his order for a contract with a market tightness $\theta_i(a_j^g)$ maximizing (7). Therefore, the inflow–outflow equations for the steady-state distribution, $n_i^g(a_j)$, of group $g$ write:

$$\text{if } i \neq j \ : 0 = \delta \sum_{k \in \mathcal{I}} n_k \left(a_j^g\right) \pi_{ki} - \delta n_i^g \left(a_j^g\right) - \alpha \left[\theta_i \left(a_j^g\right)\right] n_i \left(a_j^g\right), \tag{A4}$$

$$\text{if } i = j \ : 0 = \delta \sum_{k \in \mathcal{I}} n_k \left(a_i^g\right) \pi_{ki} - \delta n_i \left(a_j^g\right) + \sum_{k \in \mathcal{I}} \alpha \left[\theta_i \left(a_k^g\right)\right] n_k \left(a_i^g\right), \tag{A5}$$

$$\mu^g = \sum_{(i,j) \in \in \mathcal{I}^2} n_i \left(a_j^g\right). \tag{A6}$$

The first step is to establish the following result.

LEMMA A2. *The system of steady-state equations (A4)–(A6) has at least one solution.*

PROOF. The argument is standard. The system (8)–(9) characterizes the ergodic distributions of the continuous time Markov chain induced by the type switching and trading process. Denote the transition intensities by $\lambda_{k,\ell}$, and let $\lambda_k \equiv \sum_\ell \lambda_{k,\ell}$. The steady-state equations can be rewritten:

$$\forall \ell : \ \lambda_\ell n_\ell = \sum_k n_k \lambda_{k,\ell} \iff \forall \ell : \ \lambda_\ell n_\ell = \sum_k \lambda_k n_k \frac{\lambda_{k,\ell}}{\lambda_k} \iff \pi = \pi Q,$$

where $\pi_k \equiv \lambda_k n_k$ and $Q_{\ell,k} = \frac{\lambda_{\ell,k}}{\lambda_\ell}$. Thus, a steady-state distribution can be found by solving for an ergodic distribution, $\pi$, of the discrete time Markov chain with transition probabilities $Q$. Since this discrete time Markov chain has a finite state space, it follows from theorem 11.1 of Stokey and Lucas (1989) that such an ergodic distribution exists. □

Equipped with this result, we can define the aggregate demand correspondence as follows. In matrix form, the system of steady-state equations of group $g$, (A4)–(A6), can be written as $\Gamma(m^g)n^g = b$, where the entries of $\Gamma(m^g)$ are continuous in $a_i^g$ and $\theta_i(a_j^g)$. Then, we let

$$\mathcal{D}(p) \equiv \left\{ \sum_{g,i,j} \mu^g n_i \left(a_j^g\right) a_j^g, \text{ for } \mu^g \in \Delta^G, m^g \right.$$
$$\left. = (a_i, \theta_i(a_j)) \in \mathcal{M}(p), \text{ and } n^g \text{ s.t. } \Gamma(m^g)n^g = b \right\}.$$

Now let us turn to the proof of Proposition 2. By Lemma A2, the aggregate demand correspondence is nonempty. It is convex by construction. To see that it is compact-valued, consider any sequence of elements of $\mathcal{D}(p)$ generated by some sequences $\mu_k^g$, $m_k^g$, and $n_k^g$. Because group measures, asset holdings, and steady-state measures are all bounded, we can extract convergence subsequences $\mu^g_{k_\ell}$, $m^g_{k_\ell}$, and $n^g_{k_\ell}$. Since the simplex and $\mathcal{M}(p)$ are compact, it follows that $\lim \mu^g_{k_\ell} \in \Delta^G$ and $\lim m^g_{k_\ell} \in \mathcal{M}(p)$. Since $\Gamma(m)$ is continuous, it follows that $\Gamma(\lim m^g_{k_\ell}) \lim n^g_{k_\ell} = b$. Therefore, the aggregate excess demand generated by $\lim \mu^g_{k_\ell}$, $\lim m^g_{k_\ell}$, and $\lim n^g_{k_\ell}$ belongs to $\mathcal{D}(p)$, and we are done proving compactness. Finally, a similar reasoning establishes that $\mathcal{D}(p)$ is upper hemi continuous.

Now, we note that the aggregate demand goes to zero as $p$ goes to infinity, and to infinity as $p$ goes to zero. An application of the Intermediate Value Theorem (easily extended for upper hemi continuous, compact, and convex-valued correspondences) establishes the claim of the proposition.

The last step is to verify the equilibrium condition for the candidate equilibrium objects shown in the text after Proposition 2. Without loss of generality at this stage of the analysis, let us assume that $G = 1$ so that we can simplify notations and drop the $g$ subscript everywhere. We first show that $\Theta[\sigma_i(a_j)] = \theta_i(a_j)$. First, we prove that

$$V_i(a_j, \sigma_i(a_j), \theta_i(a_j)) > V_i(a_j, \mathbf{0}, 0).$$

By construction, we have a weak inequality. Suppose, toward a contradiction, that $V_i(a_j, \sigma_i(a_j), \theta_i(a_j)) = V_i(a_j, \mathbf{0}, 0)$. Expressing this equality using (A1) and the dealer's zero-profit condition, we have

$$\frac{\alpha[\theta_i(a_j)]}{\theta_i(a_j)} \left[ V_i(a + q_i(a_j)) - pq_i(a_j) - V_i(a, \mathbf{0}, 0) \right] = \gamma.$$

Now the right-derivative at $\theta = 0$ of the right-hand side of (A3) is equal to

$$\frac{\alpha'(0) \left[ V_i(a + q_i(a_j)) - pq_i(a_j) - V_i(a, 0) \right] - \gamma}{r + \delta} > 0,$$

using the equality we derived just above and noting that $\alpha'(0) > \alpha(\theta)/\theta$ by strict concavity. Hence, the right-hand side of the auxiliary Bellman equation (A3) cannot be maximized at $\theta = 0$, which is a contradiction. Thus, we are in case 2 of Lemma A1, $V_i(a_j, \sigma_i(a_j), \theta_i(a_j))$ is strictly increasing, $\theta_i(a_j) = \inf\{\theta \geq 0 \, : \, V_i(a_j, \sigma_i(a_j), \theta) > V_i^\star(a_j)\}$ and so is greater than $\Theta(\sigma_i(a_j))$.

Suppose, toward a contradiction, that the inequality is strict. Then, by definition of $\Theta(\sigma)$, there exists some $\theta \in (0, \theta_i(a_j))$ and some $(k, \ell)$ such that $V_k(a_\ell, \sigma_i(a_j), \theta) > V_k^\star(a_\ell)$. Using the definition of $V_k(a, \sigma, \theta)$, we obtain that

$$\frac{u_k(a_\ell) + \delta \sum_{m\in\mathcal{I}} \pi_{km} V_m^\star(a_\ell) + \alpha(\theta)\left[V_k^\star(a_\ell + q_i(a_j)) - pq_i(a_j)\right] - \gamma\theta \dfrac{\alpha(\theta)}{\theta} \dfrac{\theta_i(a_j)}{\alpha(\theta_i(a_j))}}{r + \delta + \alpha(\theta)} > V_k^\star(a_\ell).$$

Since $\alpha(\theta)/\theta$ is decreasing, it follows that the above inequality remains strict when we subtract $\gamma\theta$ instead of $\gamma\theta\frac{\alpha(\theta)}{\theta}\frac{\theta_i(a_j)}{\alpha[\theta_i(a_j)]}$, which contradicts that $V_k^\star(a)$ solves the auxiliary Bellman equation (A3).

Finally, we verify the zero-profit conditions of dealers. Suppose there is $\sigma = (q, \phi)$ such that $-\gamma + \frac{\alpha[\Theta(\sigma)]}{\Theta(\sigma)}\phi > 0$. Then $\Theta(\sigma) < \infty$ and there is some type $(i, a_j)$ and some tightness $\theta$ such that $V_i(a_j, \sigma, \theta) > V_i^\star(a_j)$ and $-\gamma + \frac{\alpha(\theta)}{\theta}\phi > 0$. Consider then the contract $\hat\sigma$, with $\hat q = q$ and $\hat\phi$ chosen such that $\gamma = \frac{\alpha(\theta)}{\theta}\hat\phi$. Because $\hat\phi < \phi$, we have that $V_i(a_j, \hat\sigma, \theta) > V_i^\star(a_j)$, which can be written as

$$\frac{u_i(a_j) + \delta \sum \pi_{ik} V_k^\star(a_k) + \alpha(\theta)\left[V_i^\star(a_j + \hat q) - p\hat q\right] - \gamma\theta}{r + \delta + \alpha(\theta)} > V_i^\star(a_j),$$

contradicting that $V_i^\star(a)$ solves the Bellman equation (A3). □

### *A.3 Market Clearing in the Interdealer Market*

In this section, we confirm that the equilibrium conditions described in Definition 1 ensure that the interdealer market clears, that is, that

$$\sum_{i,j} \alpha[\theta_i(a_j)] n_i(a_j)(a_i - a_j) = 0. \tag{A7}$$

To start, summing across all $i \in \mathcal{I}$ and $j \neq i$, we can use (8) to get

$$\sum_{i\in\mathcal{I}} \sum_{j\neq i} \alpha[\theta_i(a_j)] n_i(a_j) a_j = \delta \sum_{i\in\mathcal{I}} \sum_{j\neq i} a_j \left[\sum_{k\in\mathcal{I}} \pi_{ki} n_k(a_j) - n_i(a_j)\right]. \tag{A8}$$

Adopting the convention that $\theta_i(a_i) = 0$, the left-hand side of (A8) is equal to

$$\sum_{i,j} \alpha[\theta_i(a_j)] n_i(a_j) a_j. \tag{A9}$$

Meanwhile, the right-hand side of (A8) can be written as

$$\delta \sum_{k\in\mathcal{I}} \sum_{j\in\mathcal{I}} n_k(a_j) a_j \sum_{i\in\mathcal{I}} \pi_{ki} - \delta \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{I}} n_i(a_j) a_j$$
$$- \sum_{i\in\mathcal{I}} \left\{\delta \left[\sum_{k\in\mathcal{I}} n_k(a_i)\pi_{ki} - n_i(a_i)\right]\right\} a_i$$

$$= -\sum_{i \in \mathcal{I}} \left\{ \delta \left[ \sum_{k \in \mathcal{I}} n_k(a_i)\pi_{ki} - n_i(a_i) \right] \right\} a_i \tag{A10}$$

$$= \sum_{i,j} \alpha[\theta_i(a_j)] n_i(a_j) a_i, \tag{A11}$$

where the first equality follows from $\sum_i \pi_{ki} = 1$, while the second equality follows from (9). Hence, (A9) is equal to (A11), which ensures that (A7) is satisfied.

### *A.4 Proof of Lemma 1*

Let, for $i \in \{\ell, h\}$, $\theta_i(0) \equiv \arg\max_\theta\{\alpha(\theta)(\Delta V_i - p) - \gamma\theta\}$ and $\theta_i(1) \equiv \arg\max_\theta\{\alpha(\theta)(p - \Delta V_i) - \gamma\theta\}$. We obtain the inequalities:

$$r\Delta V_h \geq u_h + \delta\pi_\ell\,(\Delta V_\ell - \Delta V_h) + \alpha[\theta_\ell(1)]\,(p - \Delta V_h) - \gamma\theta_\ell(1) - \alpha[\theta_h(0)]$$
$$(\Delta V_h - p) - \gamma\theta_h(0),$$
$$r\Delta V_\ell \leq u_\ell + \delta\pi_h\,(\Delta V_h - \Delta V_\ell) + \alpha[\theta_\ell(1)]\,(p - \Delta V_\ell) - \gamma\theta_\ell(1) - \alpha[\theta_h(0)]$$
$$(\Delta V_\ell - p) - \gamma\theta_h(0).$$

Taking the difference between the two, all the search costs cancel out and we obtain

$$\Delta V_h - \Delta V_\ell \geq \frac{u_h - u_\ell}{r + \delta + \alpha\,[\theta_h(0)] + \alpha\,[\theta_\ell(1)]} > 0,$$

which establishes the first part of the claim.

For the second part, assume toward a contradiction that $p \geq \Delta V_h$, so that $p > \Delta V_\ell$. Then, it must be the case that $\theta_h(1) \geq 0$, $\theta_\ell(1) > 0$, and $\theta_h(0) = \theta_\ell(0) = 0$. The inflow–outflow equations for $n_h(1)$ and $n_\ell(1)$ become

$$\delta\pi_h n_\ell(1) = \delta\pi_\ell n_h(1) + \alpha[\theta_h(1)] n_h(1),$$
$$\delta\pi_\ell n_h(1) = \delta\pi_h n_\ell(1) + \alpha[\theta_\ell(1)] n_\ell(1),$$

implying, after some manipulations, that

$$\frac{\delta\pi_\ell}{\delta\pi_\ell + \alpha[\theta_h(1)]}\,\frac{\delta\pi_h}{\delta\pi_h + \alpha[\theta_\ell(1)]} n_\ell(1) = n_\ell(1),$$

so that $n_\ell(1) = n_h(1) = 0$. Thus, the market cannot clear, which is a contradiction. Symmetrically, if $p \leq \Delta V_\ell$, one finds that $n_\ell(0) = n_h(0) = 0$ so that $n_\ell(1) + n_h(1) = 1 > A$, which also contradicts market clearing. □

### *A.5 Proof of Lemma 2*

The function $F(S_\ell)$ is defined over $[0, \Gamma^{-1}(u_h - u_\ell)]$ and the function $G(S_\ell)$ is defined over $[0, \frac{u_h - rp}{\delta\pi_\ell}]$. Both functions are zero at the upper bound of their domain. Thus, solving the system of equations (15)–(16) boils down to solving the one-equation-in-one-unknown problem $F(S_\ell) = G(S_\ell)$. We have shown in the text that if a solution exists, then the function $F(S_\ell)$ must cross the function $G(S_\ell)$ from above. Therefore, for a strictly positive solution to exist, it is necessary and sufficient that $F(S_\ell)$ is above $G(S_\ell)$ at zero, and is eventually below $G(S_\ell)$ for $S_\ell$ large enough in the intersection of their domains. The first condition can be written

$$\begin{aligned} F(0) > G(0) &\iff \Gamma[F(0)] > u_h - rp \text{ and } \Gamma[F(0)] = u_h - u_\ell \\ &\iff \delta\pi_h F(0) < rp - u_\ell \text{ and } \Gamma[F(0)] = u_h - u_\ell \\ &\iff u_h - u_\ell < \Gamma\left(\frac{rp - u_\ell}{\delta\pi_h}\right) \\ &\iff p > p_\ell \text{ s.t. } u_h - u_\ell = \Gamma\left(\frac{rp_\ell - u_\ell}{\delta\pi_h}\right). \end{aligned} \tag{A12}$$

The equivalence on the first line follows from the fact that, in equation (18), the function $\Gamma(S)$ is strictly increasing. The equivalence on the second line follows from simple manipulations, using the definition of $\Gamma(S)$. The equivalence on the third and fourth lines follow from the fact that $\Gamma(S)$ is strictly increasing in $S$. One sees easily using equations (15) and (16) that the price $p_\ell$ is such that $S_\ell = 0$; in other words, it is the lowest price at which a low-valuation investor is willing to sell his asset, knowing that he will be able to buy at this price later when he will turn into a high type.

Next we note that, by definition, $F(S_\ell)$ and $G(S_\ell)$ are zero at the upper bound of their respective domains. Thus, in order for $G(S_\ell)$ to be below $F(S_\ell)$ for $S_\ell$ large enough in the intersection of their domains, it is necessary and sufficient that the upper bound of the domain of $G(S_\ell)$ is below the upper bound of the domain of $F(S_\ell)$. This can be written as

$$\begin{aligned} \Gamma^{-1}(u_h - u_\ell) > \frac{u_h - rp}{\delta\pi_\ell} &\iff u_h - u_\ell > \Gamma\left(\frac{u_h - rp}{\delta\pi_\ell}\right) \\ &\iff p < p_h \text{ s.t. } u_h - u_\ell = \Gamma\left(\frac{u_h - rp_h}{\delta\pi_\ell}\right). \end{aligned} \tag{A13}$$

As before, one sees easily that the price $p_h$ is such that $S_h = 0$: it is the highest price at which a high type is willing to buy.

Finally, we observe that $p_\ell < p_h$. Indeed, from their definitions we have that:

$$\frac{u_h - rp_h}{\delta\pi_\ell} = \frac{rp_\ell - u_\ell}{\delta\pi_h} \Rightarrow rp_h\pi_h + rp_\ell\pi_\ell = \pi_h u_h + \pi_\ell u_\ell.$$

It thus follows that

$$p_h > p_\ell > 0 \iff r\pi_\ell(p_h - p_\ell) > 0 \iff rp_h > \pi_h u_h + \pi_\ell u_\ell$$
$$\iff \Gamma\left(\frac{u_h - \pi_h u_h - \pi_\ell u_\ell}{\delta\pi_\ell}\right) > u_h - u_\ell$$
$$\iff \Gamma\left(\frac{u_h - u_\ell}{\delta}\right) > u_h - u_\ell.$$

One easily verifies that $\Gamma(S/\delta) > S$, so this inequality holds. □

*A.6 Proof of Proposition 3*

The only thing to show is that the buy-order flow, on the left-hand side of (25), is strictly decreasing in the price, and that the sell-order flow, on the right-hand side of (25), is strictly increasing in the price. For this we first note that, since $\alpha(\theta)$ is increasing, it follows that the left-hand side is an increasing function of $\theta_h$ and the right-hand side is a decreasing function of $\theta_\ell$. Now the first-order condition $\alpha'(\theta_h)S_h = \gamma$ implies that $\theta_h$ is an increasing function of $S_h$, and thus a decreasing function of $p$, with $\theta_h = 0$ when $p = p_h$. Similarly, the first-order condition for $\theta_\ell$ implies that $\theta_\ell$ is an increasing function of $S_\ell$ and thus an increasing function of $p$ with $\theta_\ell = 0$ when $p = p_\ell$. □

*A.7 Proof of Proposition 4*

*Asymptotic expansion for the price.* Let $s_i \equiv \lambda S_i$ for $i \in \{\ell, h\}$. With this notation, the surplus equation (17) can be written as

$$\frac{r+\delta}{\lambda}s_h + f(s_h) + \frac{r+\delta}{\lambda}s_\ell + f(s_\ell) = u_h - u_\ell,$$

where $f(s) \equiv \max_\theta\{\alpha(\theta)s - \gamma\theta\}$. Note that $f(s)$ is strictly increasing, with $f(0) = 0$ and $f(\infty) = \infty$. It thus follows that $s_h$ and $s_\ell$ are both bounded by the solution of $f(s) = u_h - u_\ell$. Therefore, as $\lambda$ goes to infinity, $(s_h, s_\ell)$ must have at least one accumulation point, $(s_h^\star, s_\ell^\star)$. Since $\pi_h > A$, it follows from the market-clearing condition that $\theta_h < \theta_\ell$ and so from the first-order condition that $s_h < s_\ell$. Thus, $s_h^\star \le s_\ell^\star$. Suppose that $s_h^\star > 0$. Then the market tightness solving $\alpha'(\theta)s_h^\star = \gamma$ is strictly positive, and going to the $\lambda \to \infty$ limit in the market-clearing condition leads to $\pi_h = A$, which contradicts our assumption that $\pi_h > A$. Therefore, $s_h^\star$ and $s_\ell^\star$ satisfy

$$s_h^\star = 0,$$
$$f\left(s_\ell^\star\right) = u_h - u_\ell.$$

Thus, $(s_h, s_\ell)$ has a unique accumulation point, which must be its limit. To obtain a first-order approximation for $p$ we use the second equation (18) to get that

$$\frac{r+\delta\pi_\ell}{\lambda}s_h + f(s_h) = u_h - rp - \delta\pi_\ell\frac{s_\ell}{\lambda}. \tag{A14}$$

Since $s_h \to 0$ as $\lambda \to \infty$, the first term on the left-hand side is $o(\frac{1}{\lambda})$. To analyze the second term, note from the market-clearing condition that $\lambda\alpha(\theta_h)$ must have a positive limit, equating order flows when sellers can contact the market infinitely fast. Letting this limit be $\lambda_h$ we can write

$$f(s_h) = \left[\frac{\lambda_h}{\lambda} + o\left(\frac{1}{\lambda}\right)\right]s_h - \gamma\alpha^{-1}\left[\frac{\lambda_h}{\lambda} + o\left(\frac{1}{\lambda}\right)\right].$$

The first term on the left-hand side is $o(\frac{1}{\lambda})$ since $s_h \to 0$ when $\lambda \to \infty$. Because $\alpha^{-1}(x)$ is convex, the second term can be bounded by

$$\begin{aligned}\alpha^{-1}\left[\frac{\lambda_h}{\lambda} + o\left(\frac{1}{\lambda}\right)\right] &\leq \left[\frac{\lambda_h}{\lambda} + o\left(\frac{1}{\lambda}\right)\right]\left[\alpha^{-1}\right]'\left[\frac{\lambda_h}{\lambda} + o\left(\frac{1}{\lambda}\right)\right] \\ &= \left[\frac{\lambda_h}{\lambda} + o\left(\frac{1}{\lambda}\right)\right]\frac{1}{\alpha'\circ\alpha^{-1}\left[\frac{\lambda_h}{\lambda} + o\left(\frac{1}{\lambda}\right)\right]}.\end{aligned}$$

Clearly, because of the Inada condition $\alpha'(0) = \infty$, the upper bound is $o(\frac{1}{\lambda})$. We conclude that $f(s_h)$ is $o(\frac{1}{\lambda})$. Finally,we note that $\frac{s_\ell}{\lambda} = \frac{s_\ell^\star}{\lambda} + o(\frac{1}{\lambda})$.

*Asymptotic expansion of the intermediation fee.* Consider first the intermediation fee for a seller:

$$\lambda\phi_\ell = \frac{\theta_\ell\alpha'(\theta_\ell)}{\alpha(\theta_\ell)}\lambda S_\ell \to \frac{\theta_\ell^\star\alpha'(\theta_\ell^\star)}{\alpha(\theta_\ell^\star)}s_\ell^\star,$$

as claimed in the proposition. Let us turn next to the intermediation fee for buyers:

$$\lambda\phi_h = \frac{\theta_h\alpha'(\theta_h)}{\alpha(\theta_h)}\lambda S_h \to 0$$

since the elasticity is bounded above by one and $\lim_{\lambda\to\infty}\lambda S_h = 0$.

*Asymptotic expansion of the distribution of types.* From equation (24), given that $\theta_\ell \to \theta_\ell^\star$, we have that

$$\lim_{\lambda\to\infty}\lambda n_\ell(1) = \frac{A\delta\pi_\ell}{\alpha(\theta_\ell^\star)}$$

as claimed. To obtain an asymptotic expansion of $n_h(0)$, plug the equality of order flows, (20), into the equation for $n_h(0)$, (23). In doing so, keep in mind that for the asymptotic expansion, all the $\alpha(\theta)$ are multiplied by the search efficiency parameter $\lambda$. We then obtain

$$n_h(0) = \frac{\delta\pi_h(1-A)}{\delta + \frac{\lambda\alpha(\theta_\ell)n_\ell(1)}{n_h(0)}} \iff \delta n_h(0) + \lambda\alpha(\theta_\ell)n_\ell(1) = (1-A)\delta\pi_h$$

$$\iff \delta\left[n_h(0) - (\pi_h - A)\right] = -\lambda\alpha(\theta_\ell)n_\ell(1) + (1-A)\delta\pi_h - \delta(\pi_h - A)$$

$$\iff n_h(0) - (\pi_h - A) = \frac{\delta A\pi_\ell}{\lambda\alpha(\theta_\ell)} = \frac{\delta A\pi_\ell}{\lambda\alpha(\theta_\ell^\star)} + o\left(\frac{1}{\lambda}\right),$$

where the last line follows after substituting in expression (24) for $n_\ell(1)$. This establishes the claim. □

### A.8 *Proof of Corollary 1*

With a Cobb–Douglas matching function, the asymptotic seller surplus, $s_\ell^\star$, and the asymptotic search intensity, $\theta_\ell^\star$, solve the system of equations:

$$\theta^\eta s - \gamma\theta = u_h - u_\ell,$$
$$\eta\theta^{\eta-1}s = \gamma.$$

The second equation implies that $\eta\theta^\eta s_\ell^\star = \gamma\theta$. Together with the first equation, this implies that $(1-\eta)\theta^\eta s = u_h - u_\ell$. Dividing through by the second equation, the $s$ cancels out and we obtain the expression for $\theta_\ell^\star$. The expression for $s_\ell^\star$ follows. Clearly, $\theta_\ell^\star$ is increasing in $\eta$. As for $s_\ell^\star$ we have:

$$\log\left(s_\ell^\star\right) = \eta\log\left(\frac{\gamma}{\eta}\right) + (1-\eta)\log\left(\frac{u_h - u_\ell}{1-\eta}\right)$$

$$\Longrightarrow \frac{d\log(s_\ell^\star)}{d\eta} = \log\left(\frac{\gamma(1-\eta)}{\eta(u_h - u_\ell)}\right),$$

which is strictly decreasing, goes to plus infinity when $\eta \to 0$ and to minus infinity when $\eta \to 1$. □

### A.9 *Proof of Proposition 5*

Let us denote $\vartheta_i(a)$ the current-value costate variables associated with $n_i(a)$ and $\xi$ the Lagrange multiplier associated with the balanced order flow constraint. The current-value Hamiltonian is

$$\mathcal{H}\left[\theta_h, \theta_\ell, n_i(a), \vartheta_i(a), \xi\right] = n_h(1)u_h + n_\ell(1)u_\ell - \theta_h n_h(0)\gamma - \theta_\ell n_\ell(1)\gamma$$
$$+\vartheta_h(1)\left\{\alpha(\theta_h)n_h(0) + \delta\pi_h n_\ell(1) - \delta\pi_\ell n_h(1)\right\}$$
$$+\vartheta_h(0)\left\{\delta\pi_h n_\ell(0) - \delta\pi_\ell n_h(0) - \alpha(\theta_h)n_h(0)\right\}$$

$$+\vartheta_\ell(1)\{\delta\pi_\ell n_h(1)-\delta\pi_h n_\ell(1)-\alpha(\theta_\ell)n_\ell(1)\}$$
$$+\vartheta_\ell(0)\{\alpha(\theta_\ell)n_\ell(1)+\delta\pi_\ell n_h(0)-\delta\pi_h n_\ell(0)\}$$
$$+\xi\,\{\alpha(\theta_\ell)n_\ell(1)-\alpha(\theta_h)n_h(0)\}.$$

The first-order conditions for $\theta_h$ and $\theta_\ell$ are

$$\alpha'(\theta_h)\left[\vartheta_h(1)-\vartheta_h(0)-\xi\right]=\gamma, \tag{A15}$$

$$\alpha'(\theta_\ell)\left[\vartheta_\ell(0)-\vartheta_\ell(1)+\xi\right]=\gamma. \tag{A16}$$

The social cost of assigning a dealer to the buy side of the market corresponds to the entry cost, $\gamma$, the right-hand side of (A15). The social benefit, the left-hand side of (A15), is measured by the increase in the number of matches generated by a dealer, $\alpha'(\theta_h)$, times the social gain from having an additional $h$-type investor holding an asset, $\vartheta_h(1)-\vartheta_h(0)$, net of the cost of tightening the balanced order flow constraint, $\xi$. The ODEs for the costate variables are

$$r\vartheta_h(1)=u_h+\delta\pi_\ell\left[\vartheta_\ell(1)-\vartheta_h(1)\right]+\dot{\vartheta}_h(1), \tag{A17}$$

$$\begin{aligned} r\vartheta_h(0)=&\ \alpha(\theta_h)\left[\vartheta_h(1)-\vartheta_h(0)-\xi\right]\\ &-\theta_h\gamma+\delta\pi_\ell\left[\vartheta_\ell(0)-\vartheta_h(0)\right]+\dot{\vartheta}_h(0), \end{aligned} \tag{A18}$$

$$\begin{aligned} r\vartheta_\ell(1)=&\ u_\ell+\alpha(\theta_\ell)\left[\vartheta_\ell(0)-\vartheta_\ell(1)+\xi\right]\\ &-\theta_\ell\gamma+\delta\pi_h\left[\vartheta_h(1)-\vartheta_\ell(1)\right]+\dot{\vartheta}_\ell(1), \end{aligned} \tag{A19}$$

$$r\vartheta_\ell(0)=\delta\pi_h\left[\vartheta_h(0)-\vartheta_\ell(0)\right]+\dot{\vartheta}_\ell(0). \tag{A20}$$

Equations (A17)–(A20) admit a standard interpretation as asset pricing conditions for the shadow values of the state variables.

It is easy to check that the planner's optimality conditions coincide with the conditions for a competitive search equilibrium when $\vartheta_i(a)=V_i(a)$ and $\xi=p$. Using that the maximized Hamiltonian is linear in the state variables, $n_i(a)$, the equilibrium path satisfies necessary and sufficient conditions (see Arrow sufficiency theorem) for a constrained-optimal allocation. □

### *A.10 Proof of Corollary 2*

We indicate all equilibrium objects by a subscript $o$ (for opaque). The free-entry conditions for dealers are

$$\frac{\alpha(\theta_h^o)}{\theta_h^o}\nu\left[V_h^o(1)-V_h^o(0)-p^o\right]=\gamma, \tag{A21}$$

$$\frac{\alpha(\theta_\ell^o)}{\theta_\ell^o}\nu\left[V_\ell^o(0) - V_\ell^o(1) + p^o\right] = \gamma. \tag{A22}$$

The left-hand sides of (A21) and (A22) are the expected flow profits of dealers: a dealer receives an order to execute with Poisson arrival rate $\alpha(\theta_i^o)/\theta_i^o$, in which case he is paid an intermediation fee equal to a fraction $\nu$ of the match surplus as measured by $V_h^o(1) - V_h^o(0) - p^o$ if the investor is a buyer and $V_\ell^o(0) - V_\ell^o(1) + p^o$ if the investor is a seller. The investors' value functions in a steady state solve the following Bellman equations:

$$rV_h^o(1) = u_h + \delta\pi_\ell\left[V_\ell^o(1) - V_h^o(1)\right], \tag{A23}$$

$$rV_h^o(0) = \alpha\left(\theta_h^o\right)(1-\nu)\left[V_h^o(1) - V_h^o(0) - p^o\right] + \delta\pi_\ell\left[V_\ell^o(0) - V_h^o(0)\right], \tag{A24}$$

$$rV_\ell^o(1) = u_\ell + \alpha\left(\theta_\ell^o\right)(1-\nu)\left[V_\ell(0) - V_\ell(1) + p^o\right] + \delta\pi_h\left[V_h^o(1) - V_\ell^o(1)\right], \tag{A25}$$

$$rV_\ell^o(0) = \delta\pi_h\left[V_h^o(0) - V_\ell^o(0)\right]. \tag{A26}$$

In (A23) and (A26) investors hold their desired portfolio and hence the only transition that can happen is when a preference shock hits with arrival rate $\delta\pi_i$. In contrast, in (A24) and (A25) investors are able rebalance their portfolio with Poisson arrival rate $\alpha(\theta_i^o)$, in which case they enjoy a fraction $1-\nu$ of the surplus of a match with a dealer. The result follows by comparing the equilibrium conditions, (A21) and (A22) and (A23)–(A26), with the conditions for a constrained-efficient allocation, (A15) and (A16) and (A17)–(A20). □

### *A.11 Formal Proofs of the Statements in Section 3.2*

In this section, we characterize an equilibrium for the economy described in Section 3.2, when the posting cost, $\gamma$, is small.

*Equilibrium price.* We guess and verify that, in this case, the equilibrium price in LR is an equilibrium price in our environment as well.

Let $V_i^\star(a, \gamma)$ be the solution of the Bellman equation (7) when the contract posting cost is $\gamma$ and the price is equal to the equilibrium price in LR. By an application of the contraction mapping theorem, this function is continuous in $(a, \gamma)$. In particular, $V_i^\star(a, 0)$ coincides with the value function in LR. We know the following result from their paper.

LEMMA A3. *In LR, the value net of purchasing cost, $V_i^\star(a, 0) - pa$, is an increasing and affine transformation of*

$$K_i(a) = \frac{(r+\mu)u_i(a) + \delta\sum_{j\in\mathcal{I}}\pi_j u_j(a)}{r+\mu+\delta} - rpa. \tag{A27}$$

Note that $K_i(a)$ is strictly concave and it achieves its unique maximum at the asset holding prevailing in LR's equilibrium.

$$N_i(\gamma_1, \gamma_2) \equiv \left\{ a \,:\, \max_{a'} \left\{ V_i^\star(a', \gamma_1) - pa' \right\} - \frac{\gamma_2}{\mu} < V_i^\star(a, \gamma_1) - pa \right\}.$$

LEMMA A4. *The inaction regions $N_i(\gamma, \gamma)$ are disjoint for all $\gamma$ small enough.*

PROOF. Note that LR's net values, $V_i^\star(a, 0) - pa$, are strictly concave so that the inaction regions $N_i(0, \gamma_2)$ are open intervals around the optimal asset holding, $\arg\max_{a'} V_i^\star(a', 0) - pa'$, shrinking monotonically to LR's optimal asset holding when $\gamma_2 \to 0$. Consider then some $\gamma_2$ and $\varepsilon$ small enough so that all the $N_i\,(0, \gamma_2 + \mu\varepsilon)$ are disjoint. Noting that $\max_{a'}\{V_i^\star(a', \gamma) - pa'\} - V_i^\star(a, \gamma) + pa$ is continuous in $(a, \gamma)$, there exists some $\gamma_1$ such that for all $\gamma_1' \leq \gamma_1$ and all $a \in [\underline{a}, \overline{a}]$:

$$\max_{a'} \left\{ V_i^\star(a', \gamma_1') - pa' \right\} - V_i^\star(a, \gamma_1') + pa >$$
$$\max_{a'} \left\{ V_i^\star(a', 0) - pa' \right\} - V_i^\star(a, 0) + pa - \varepsilon.$$

In particular, for $a \notin N_i(0, \gamma_2 + \mu\varepsilon)$, we have $\max_{a'}\{V_i^\star(a', 0) - pa'\} - \frac{\gamma_2}{\mu} - \varepsilon \leq V_i^\star(a, 0) - pa$. Plugging this into the right-hand side of the above inequality and rearranging, we obtain

$$\max_{a'}\{V_i(a', \gamma_1') - pa'\} - \frac{\gamma_2}{\mu} \geq V_i(a, \gamma') - pa.$$

Hence, $a \notin N_i(\gamma_1', \gamma_2)$. By contrapositive, for all $\gamma_1' \leq \gamma_1$, we have that $N_i(\gamma_1', \gamma_2) \subseteq N_i(0, \gamma_2 + \mu\varepsilon)$. Since the sets $N_i(0, \gamma_2 + \mu\varepsilon)$ are disjoint by construction, and since the sets $N_i(\gamma_1', \gamma_2')$ are decreasing in $\gamma_2'$, it follows that $N_i(\gamma_1', \gamma_2')$ are disjoint for all $\gamma_1' \leq \gamma_1$ and $\gamma_2' \leq \gamma_2$. Letting $\gamma = \min\{\gamma_1, \gamma_2\}$ we obtain that the sets $N_i(\gamma, \gamma)$ are disjoint for all $\gamma' \leq \gamma$.

Next, following the argument in the text, we find that when $\gamma$ is small enough, the net value $V_i(a) - pa$ is an increasing affine transformation of $K_i(a)$. Since the optimal asset holding must belong to the *interior* of the inaction region, it must satisfy $K_i'(a) = 0$, which implies in turn that it must coincide with LR's optimal asset holdings. Since the distribution of type must also coincide with that of LR, which satisfies market clearing by construction, it follows that LR's price is an equilibrium price in our economy as long as $\gamma$ is small enough.

*Constructing a competitive search equilibrium.* The last thing to do is to verify that the LR price, which we already know clears the market, is the basis of a competitive search equilibrium. We note that for $\gamma$ small enough, there are always strict gains from trade when $i \neq j$ and so $\theta_i(a_j) = 1$. The equilibrium objects are then defined in a manner similar to the paragraph following Proposition 2:

$$q_i(a_j) = a_i - a_j, \quad \phi_i(a_j) = \frac{\gamma}{\mu}, \text{ and } \quad \theta_i(a_j) = 1.$$

We let

$$V_i^\star(a) = \max_{k,\ell} V_i \left[a, \sigma_k(a_\ell), \theta_k(a_\ell)\right].$$

We will adopt a slightly different definition for market tightness:

$$\Theta\left(\sigma\right) = \inf\left\{\theta \geq 0 \,:\, V_i(a,\sigma,\theta) \geq V_i^\star(a) \text{ and } V_i^\star(a) > V_i(a,\mathbf{0},0) \text{ for some } i \in \mathcal{I}\right\}.$$

In the case when $\alpha(\theta)$ is strictly increasing and strictly concave, this definition is equivalent to the one we used before. It is stronger in the present Leontief case.

We proceed to verify that these candidate equilibrium objects form a competitive search equilibrium. We first show that $\Theta[\sigma_i(a_j)] = \theta_i(a_j)$. First, we note that there are strict gains from trade in LR so, by the continuity arguments already used above, there must be strict gains from trade in our setup when $\gamma$ is small enough. That is,

$$V_i(a_j, \sigma_i(a_j), \theta_i(a_j)) > V_i(a_j, \mathbf{0}, 0).$$

Given that $V_i(a_j, \sigma_i(a_j), \theta_i(a_j)) = V_i^\star(a_j)$, the definition of $\Theta(\sigma)$ implies that $\Theta[\sigma_i(a_j)] \leq \theta_i(a_j)$. Suppose, toward a contradiction, that the inequality is strict. Then, by definition of $\Theta(\sigma)$, there exists some $\theta \in (0, \theta_i(a_j))$ and some $(k, \ell)$ such that $V_k(a_\ell, \sigma_i(a_j), \theta) \geq V_k^\star(a_\ell) > V_i(a_\ell, \mathbf{0}, 0)$. Since $V_k(a_\ell, \sigma_i(a_j), \theta) > V_i(a_\ell, \mathbf{0}, 0)$, one sees easily that $\theta \mapsto V_k(a_\ell, \sigma_i(a_j), \theta)$ is strictly increasing in $\theta \in (0, \theta_i(a_j))$. This implies that $V_k(a_\ell, \sigma_i(a_j), \theta_i(a_j)) > V_k^\star(a_\ell)$, which is a contradiction since $V_k^\star(a_\ell)$ is the maximum attainable utility for this investor.

Finally, we verify the zero-profit conditions of dealers. For this, we first note that in the Leontief case, the alternative representation of Proposition 1 continues to hold (one can verify that proof goes through almost identically), and so $V_i^\star(a_j)$ solves the Bellman equation (A3). Now suppose that there is $\sigma = (q, \phi)$ such that $-\gamma + \frac{\alpha[\Theta(\sigma)]}{\Theta(\sigma)}\phi > 0$. Then $\Theta(\sigma) < \infty$ and there is some type $(i, a_j)$ and some tightness $\theta$ such that $V_i(a_j, \sigma, \theta) \geq V_i^\star(a_j) > V_i(a, \mathbf{0}, 0)$ and $-\gamma + \frac{\alpha(\theta)}{\theta}\phi > 0$. Consider then the contract $\hat{\sigma}$, with $\hat{q} = q$ and $\hat{\phi}$ chosen such that $\gamma = \frac{\alpha(\theta)}{\theta}\hat{\phi}$. Because $\hat{\phi} < \phi$, we have that $V_i(a_j, \hat{\sigma}, \theta) > V_i^\star(a_j)$, which can be written

$$\frac{u_i(a_j) + \delta \sum \pi_{ik} V_k^\star(a_k) + \alpha(\theta)\left[V_i^\star(a_j + \hat{q}) - p\hat{q}\right] - \gamma\theta}{r + \delta + \alpha(\theta)} > V_i^\star(a_j),$$

contradicting that $V_i^\star(a_j)$ solves the Bellman equation (A3). □

## LITERATURE CITED

Albrecht, James, Pieter Gautier, and Susan Vroman. (2013) "Directed Search in the Housing Market." Working Paper, Georgetown University.

Boehmer, Ekkehart. (2005) "Dimensions of Execution Quality: Recent Evidence for U.S. Equity Markets." *Journal of Financial Economics*, 78, 553–82.

Burdett, Kenneth, and Maureen O'Hara. (1987) "Building Blocks: An Introduction to Block Trading." *Journal of Banking and Finance*, 11, 193–212.

Burdett, Kenneth, Shouyong Shi, and Randall Wright. (2001) "Pricing and Matching with Frictions." *Journal of Political Economy*, 109, 1060–85.

Chang, Briana. (2012) "Adverse Selection and Liquidity Distortion." Working Paper, University of Wisconsin, Madison.

Diaz, Antonia, and Belén Jerez. (2013) "House Prices, Sales, and Time on the Market: A Search-Theoretic Framework." Working Paper, Universidad de Carlos III.

Duffie, Darrell. (2012) *Dark Markets: Asset Pricing and Information Transmission in Over-the-Counter Markets*. Princeton, NJ: Princeton University Press.

Duffie, Darrell, Nicolae Gârleanu, and Lasse H. Pedersen. (2005) "Over-the-Counter Markets." *Econometrica*, 73, 1815–47.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Pedersen. (2007) "Valuation in Over-the-Counter Markets." *Review of Financial Studies*, 20, 1865–1900.

Edwards, Amy K., Lawrence E. Harris, and Michael S. Piwowar. (2007) "Corporate Bond Market Transaction Costs and Transparency." *Journal of Finance*, 62, 1421–51.

Faig, Miquel, and Belén Jerez. (2006) "Inflation, Prices, and Information in Competitive Search." *BE Journal of Macroeconomics (Advances)*.

Feldhütter, Peter. (2012) "The Same Bond at Different Prices: Identifying Search Frictions and Selling Pressures." *Review of Financial Studies*, 25, 1155–1206.

Garleanu, Nicolae.(2009) "Portfolio Choice and Pricing in Illiquid Markets." *Journal of Economic Theory*, 144, 532–64.

Gavazza, Alessandro. (2011) "Leasing and Secondary Markets: Theory and Evidence from Commercial Aircraft." *Journal of Political Economy*, 119, 325–77.

Green, Richard C., Burton Hollifield, and Norman Schürhoff. (2007) "Financial Intermediation and the Costs of Trading in an Opaque Market." *Review of Financial Studies*, 20, 275–314.

Guerrieri, Veronica, and Robert Shimer. (2012) "Dynamic Adverse Selection: A Theory of Illiquidity, Fire Sales, and Flight to Quality." Working Paper, University of Chicago.

Guerrieri, Veronica, Robert Shimer, and Randall Wright. (2010) "Adverse Selection in Competitive Search Equilibrium." *Econometrica*, 78, 1823–62.

Hall, George, and John Rust. (2003) "Middlemen versus Market Makers: A Theory of Competitive Exchange." *Journal of Political Economy*, 111, 353–403.

Harris, Lawrence E., and Michael S. Piwowar. (2004) "Secondary Trading Costs in the Municipal Bond Market." USC FBE Finance Seminar.

Hendershott, Terrence, and Ananth Madhavan. (2015) "Click or Call? Auction versus Search in the Over-the-Counter Market." *Journal of Finance*, 70, 419-47.

Ho, Thomas S. Y., and Hans R. Stoll. (1983) "The Dynamics of Dealer Markets Under Competition." *Journal of Finance*, 38, 1053–74.

Hollifield, Burton, Artem Neklyudov, and Chester Spatt. (2012) "bid–ask Spreads and the Pricing of Securitizations: 144a vs. Technical Report, Registered Securitizations." Working Paper, Carnegie Mellon University.

Hosios, Arthur J. (1990) "On the Efficiency of Matching and Related Models of Search and Unemployment." *Review of Economic Studies*, 57, 279–98.

Inderst, Roman, and Holger M. Müller. (2002) "Competitive Search Markets for Durable Goods." *Economic Theory*, 19, 599–622.

Lagos, Ricardo, and Guillaume Rocheteau. (2004) "Inflation, Output, and Welfare." Federal Reserve Bank of Cleveland Working Paper 04-07.

Lagos, Ricardo, and Guillaume Rocheteau. (2007) "Search in Asset Markets: Market Structure, Liquidity, and Welfare." *American Economic Review*, 97, 198–202.

Lagos, Ricardo, and Guillaume Rocheteau. (2009) "Liquidity in Asset Markets with Search Frictions." *Econometrica*, 77, 403–26.

Lagos, Ricardo, Guillaume Rocheteau, and Pierre-Olivier Weill. (2011) "Crises and Liquidity in Over-the-Counter Markets." *Journal of Economic Theory*, 146, 2169–2205.

Lester, Benjamin. (2011) "Information and Prices with Capacity Constraints." *American Economic Review*, 101, 1591–1600.

Lester, Benjamin, Ludo Visschers, and Ronald Wolthoff. (2013) "Competing with Asking Prices." Federal Reserve Bank of Philadelphia Working Paper Series.

Li, Dan, and Norman Schürhoff. (2012) "Dealer Networks." Working Paper, HEC Lausanne.

Lo, Andrew W., Harry Mamaysky, and Jiang Wang. (2004) "Asset Prices and Trading Volume Under Fixed Transactions Costs." *Journal of Political Economy*, 112,1054–90.

Melin, Lionel. (2012) "Multi-Speed Markets." Working Paper, University of Chicago.

Menzio, Guido, and Shouyong Shi. (2011) "Efficient Search on the Job and the Business Cycle." *Journal of Political Economy*, 119, 468–510.

Moen, Espen. (1997) "Competitive Search Equilibrium." *Journal of Political Economy*, 105, 385–411.

Mortensen, Dale T., and Randall Wright. (2002) "Competitive Pricing and Efficiency in Search Equilibrium." *International Economic Review*, 43, 1–20.

Municipal Securities Rulemaking Board. (2008) *Fact Book*.

Neklyudov, Artem. (2012) "bid–ask Spreads and the Over-the-Counter Interdealer Markets: Core and Peripheral Dealers." Working Paper, HEC Lausanne.

Pagnotta, Emiliano, and Thomas Philippon. (2012) "Competing on Speed." Working Paper, NYU Stern School of Business.

Pissarides, Christopher A. (2000) *Equilibrium Unemployment Theory*, Vol. 1. Cambridge, MA: MIT Press.

Praz, Remy. (2012) "Equilibrium Asset Pricing and Portfolio Choice in the Presence of Both Liquid and Illiquid Markets." Working Paper, EPFL.

Rocheteau, Guillaume, and Pierre-Olivier Weill. (2011) "Liquidity in Frictional Asset Markets." *Journal of Money, Credit and Banking*, 43, 261–82.

Rocheteau, Guillaume, and Randall Wright. (2005) "Money in Search Equilibrium, in Competitive Equilibrium, and in Competitive Search Equilibrium." *Econometrica*, 73, 175–202.

Rubinstein, Ariel, and Asher Wolinsky. (1987) "Middlemen." *Quarterly Journal of Economics*, 102, 581–94.

Sattinger, Michael. (2010) "Queueing and Searching." Working paper, SUNY Albany.

Sattinger, Michael. (2003) "A Search Version of the Roy Model." Working Paper, SUNY Albany.

Paul, Schultz. (2001) "Corporate Bond Trading Costs: A Peak Behind the Curtain." *Journal of Finance*, 56, 677–98.

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

Shimer, Robert. (2005) "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review*, 95, 25–49.

Spulber, Daniel F. (1996) "Market Making by Price-Setting Firms." *Review of Economic Studies*, 63, 559–80.

Stacey, Derek G. (2012) "Information, Commitment, and Separation in Illiquid Housing Markets." Working Paper, Queen's Economics Department.

Stevens, Margaret. (2007) "New Microfoundations for the Aggregate Matching Function." *International Economic Review*, 48, 847–68.

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

Vayanos, Dimitri, and Pierre-Olivier Weill. (2008) "A Search-Based Theory of the On-the-Run Phenomenon." *Journal of Finance*, 63, 1361–98.

Watanabe, Makoto. (2013) "Middlemen: A Directed Search Equilibrium Approach." Working Paper, VU University Amsterdam, Tinbergen Institute.

Weill, Pierre-Olivier. (2007) "Leaning against the Wind." *Review of Economic Studies*, 74, 1329–54.

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

Wong,Tsz-Nga. (2013) "A Tractable Monetary Model under General Preferences." Working Paper, Bank of Canada.

Zhang, Shengxing. (2012) "Liquidity Missallocation in an Over-the-Counter Market." Working Paper, NYU.