Intermediation in Over-the-Counter Markets with Price Transparency*

Miroslav Gabrovski† [1] and Ioannis Kospentaris‡[2]

[1]University of Hawaii Manoa
[2]Virginia Commonwealth University

September 22, 2021

Abstract

A salient feature of over-the-counter (OTC) markets is intermediation: dealers buy from and sell to customers as well as other dealers. Traditionally, the search-theoretic literature of OTC markets has rationalized this as a consequence of random meetings and ex post bargaining between investors. We show that neither of these are necessary conditions for intermediation. We build a model of a fully decentralized OTC market in which search is directed and sellers post prices ex ante. Intermediation arises naturally as an equilibrium outcome for a broad class of matching functions commonly used in the literature. We further explore, both analytically and numerically, how the extent of intermediation depends on the nature of frictions and model primitives. Our numerical exercises also contrast the model's equilibrium implications to those of a benchmark model with random meetings and ex post bargaining.

JEL Classification: G11, G12, G21, D83

Keywords: Over-the-counter markets, search frictions, intermediation, price transparency, competitive search, sorting

* We are grateful to Zach Bethune, Hector Chade, Briana Chang, Michael Choi, Athanasios Geromichalos, Paul Jackson, Janet Jiang, Chris Herrington, Oleg Korenok, Benjamin Lester, Florian Madison, Joe Ostroy, Roberto Pinheiro, Guillaume Rocheteau, Carol Scotese, Mario Silva, Bruno Sultanum, Semih Üslü, Liang Wang, Pierre-Olivier Weill, as well as seminar participants in University of Hawaii-Manoa, UCI, UCLA, VCU, and Search & Matching Virtual Brown Bag for helpful comments and suggestions.

† Department of Economics, University of Hawaii-Manoa; email: mgabr@hawaii.edu

‡ Department of Economics, Virginia Commonwealth University; email: ikospentaris@vcu.edu

1 Introduction

A large number of securities, including government, municipal, and corporate bonds; asset-backed securities; and federal funds, are traded in decentralized or "over-the-counter" (OTC) markets. A prevalent feature of OTC markets is *intermediation*: trades are intermediated by dealers who buy from and sell to customers, as well as other dealers.[1] Traditionally, the theoretical literature building off of the influential work of Duffie et al. (2005, 2007) has rationalized extensive intermediation as a mechanical consequence of the OTC markets' decentralized nature (finding a counterparty to trade takes time) and opaqueness (the trading parties negotiate the terms of trade privately). In recent years, however, OTC markets have become far more transparent due to the growing use of electronic trading platforms allowing traders at once to obtain quotes at which many dealers are willing to trade.[2] Yet, intermediation remains a key feature of OTC markets, as shown by the recent empirical evidence in Adrian et al. (2017) and Bessembinder et al. (2020). This presents a challenge for the traditional mechanism behind intermediation found in the theoretical literature.

To address this challenge, we develop a fully decentralized search-theoretic model of an OTC market with price transparency.[3] In the model, heterogeneous asset owners post publicly available prices at which they are willing to sell the asset and heterogeneous non-owners choose to which price to direct their orders. Both asset prices and contact rates between investors are endogenous and reflect investors' heterogeneous liquidity needs. We show that intermediation arises naturally in this framework where agents *choose* their trading counterparties based on available prices. We also explore how the extent of intermediation

[1] The literature has proposed several definitions of intermediation. We follow Hugonnier et al. (2018, 2021) and say an agent acts as an intermediary if she is actively trying to buy the asset when she does not have it and actively trying to sell it when she does. Empirical evidence for intermedation in OTC markets abounds; see Bessembinder and Maxwell (2008), Afonso and Lagos (2014), and Li and Schürhoff (2019).

[2] For evidence on electronic platforms, see Stafford (2016), Liu et al. (2018), and Vogel (2019).

[3] We use "price transparency" to capture the idea that prices are available before trading and contrast it with the ex post bargaining, commonly assumed in the search-theoretic OTC literature. In this context, it is synonymous to "price competition" from the competitive search literature. We use the two interchangeably.

depends on model primitives. Our numerical exercises contrast the model's equilibrium implications to those of a benchmark model with random meetings and ex post bargaining.

The environment follows the literature initiated by Duffie et al. (2005, 2007). The economy is populated by infinitely-lived investors who are either owners or non-owners of a single indivisible asset in fixed supply. Investors are heterogeneous in their valuation of the asset. There are finitely many valuations and we refer to each as an investor's type. Periodic preference shocks change investors' types and this creates incentives to trade. Agents can hold either zero or one unit of the asset. There are two major points of departure from Duffie et al. (2005): the asset market is fully decentralized and there is price competition. The first one allows us to study the trading pattern of the asset market as an endogenous equilibrium outcome. As in Hugonnier et al. (2021), all trades in the economy take place in bilateral meetings without a perfectly competitive dealer market. Thus, the roles of agents as either customers (those who only buy or sell the asset) or dealers (those who intermediate trades between other agents) are endogenous in our model. The second departure, namely competitive search, allows us to incorporate the empirically observed price transparency in our model. Specifically, owners post prices at which they are willing to sell the asset. Non-owners observe all posted prices and choose to direct their search towards a specific price.

Investors self-select into submarkets with endogenous meeting rates.[4] Even though buyers observe prices, there are still frictions which preclude immediate trade execution, due to institutional, informational, or technological constraints. As in Lagos and Rocheteau (2007, 2009) and Lester et al. (2015), we capture these frictions with a matching function — the execution time is exponentially distributed with mean governed by the submarket queue length (the ratio of buyers to sellers). The higher the queue length, the higher (lower) the order filling rate for sellers (buyers). Thus, submarkets are characterized by a price and a

[4] We refer to the collection of all owners who post a particular price and all non-owners who direct their orders towards that same price as a "submarket". Owners and non-owners who participate in a submarket are called sellers and buyers.

contact rate, both of which are endogenous. As is standard in models of competitive search, the collection of submarkets that open in equilibrium coincides with the planner's allocation.

Our main result is that intermediation arises naturally, even though agents *choose* their trading counterparties. Intuitively, consider the planner's problem with three investor types (low, medium, and high valuation). Suppose the low type is an owner, whereas the medium and high types are non-owners. Then, welfare will improve if the asset reaches the high type investor. The planner can instruct the low and high type agents to trade directly or the medium type to act as an intermediary: receive the asset from the low and transfer it to the high type. Due to search frictions, however, there are trading delays every time agents transfer the asset. Thus, intermediation comes at a cost: the asset generally takes longer to reach the high type non-owner. But it also comes with a benefit: the low type owner transfers the asset fast to the medium type, who receives higher flow utility while waiting to hand over the asset to the high type.

The planner finds intermediation optimal if and only if the benefit from the medium type holding the asset outweighs the cost of the extra time needed for the asset to reach the high type. In particular, if she can assign a large enough queue length of medium type non-owners to trade with low type owners, this will result in little trading delays for the low type and low intermediation costs. We show that when the matching function satisfies an asymptotic condition, commonly assumed in the literature, these costs become negligent.[5] As a result, intermediation is always optimal, regardless of how small the difference in valuations between the medium and high types is. In the case with many investor types, it is optimal for all except the lowest and highest types to act as intermediaries; i.e. intermediation is ubiquitous.

If the matching function does not satisfy the condition, intermediation is still likely to occur but to a limited extent, because its cost is bounded away from zero. We analyti-

[5] This asymptotic condition imposes an Inada-type restriction on the matching function: the matching speed for sellers (buyers) approaches infinity as the queue length goes to infinity (zero).

cally explore how the extent of intermediation depends on model primitives through several illustrative examples. Interestingly, intermediation never takes place when the matching technology is linear. When only one side of the market contributes to congestion, there is no benefit to intermediation: low type customers hold the asset just as long whether or not they trade with an intermediary. Thus, the capacity of agents to act as dealers depends on the nature of frictions: it emerges when both sides of the market contribute to congestion.

Lastly, we investigate the numerical implications of price competition in the search-theoretic OTC framework. We find that the trade-off between prices and trading speeds, which is a feature of competitive but not of random search, has important implications for two key equilibrium objects: the cost of immediacy and intermediated trade volume. In general, the cost of immediacy is more responsive to parameter changes in the random than the competitive search model, whereas the opposite is true for intermediated volume. Furthermore, the cost of immediacy can be negative with random search, but not with competitive search.

Related Literature. This paper contributes to the search-theoretic literature on OTC markets initiated by Duffie et al. (2005, 2007); for recent surveys, see Lagos et al. (2017) and Weill (2020).[6] Most closely related to ours are the papers which feature a fully decentralized asset market including Afonso and Lagos (2015), Geromichalos and Herrenbrueck (2016), Neklyudov (2019), Üslü (2019), and Hugonnier et al. (2021). The major difference between these contributions and our work is that intermediation in these papers is a mechanical consequence of random meetings and price opaqueness, whereas in ours intermediation is the result of price transparency and investors' optimal choices of trading counterparties.

A body of work explains intermediation patterns by introducing additional heterogeneity in the search-theoretic OTC framework with price opaqueness.[7] Farboodi et al. (2017a)

[6] To a lesser extent, our paper is related to the search-theoretic literature on housing markets which operate over-the-counter and assets are discrete. See, for example, Wheaton (1990), Head et al. (2014), Gabrovski and Ortego-Marti (2019, 2021a,b), Albrecht et al. (2016), and Garriga and Hedlund (2020).

[7] To be precise, these papers aim to explain the "core-periphery" structure found in many real-world

consider heterogeneity in bargaining power, whereas Farboodi et al. (2017b) consider heterogeneity in contact rates. In Bethune et al. (2019), investors are heterogeneous in their ability to learn the private valuation of the asset held by others (screening ability). In contrast to all these papers, ours features price transparency and directed search, maintaining traditional differences in investor valuations as the only source of heterogeneity. Our results imply that neither random search nor price opaqueness, nor additional heterogeneity are necessary conditions for intermediation to arise in this class of models.

Finally, our work is part of the voluminous literature on competitive search; see Wright et al. (2019) for a survey. Within that literature, the most closely related papers are the ones which feature two-sided heterogeneity such as Shi (2001), Shimer (2005), Eeckhout and Kircher (2010), and Jerez (2014). Importantly, the focus here is on OTC asset markets where agents switch roles after trading: owners become non-owners and vice versa. This opens the door to a new economic interpretation of the equilibrium assignment through *intermediation*: each trade forms a link in an *intermediation chain*, which is at the center of our analysis. In the OTC context, Lester et al. (2015) provide a detailed analysis of a semi-centralized model with a frictionless dealer market. In contrast, we consider a fully decentralized OTC market.[8] The literature, with the exception of Lester et al. (2015) and Guerrieri and Shimer (2014), has examined competitive search problems with exogenously fixed distributions of buyers and sellers. This is in contrast to our model in which the distributions are endogenous.

OTC markets, a research question outside the scope of this paper. Chang and Zhang (2018) build a bilateral matching model that combines elements of search theory with network models. In their model, agents choose whom to meet with and some emerge as intermediaries to form the core of the trading network. Our result is similar in spirit, but we derive it in the context of the workhorse search and matching environment initiated by Duffie et al. (2005).Thus, we show that one does not have to resort to alternative ways of modeling frictions to obtain intermediation when agents choose their trading counterparties.

[8] Of relevance are also Armenter and Lester (2017) who employ competitive search in a two-period model of the federal funds market with one-sided heterogeneity. A series of papers use competitive search to analyze issues of asymmetric information and asset pricing. These include Guerrieri et al. (2010), Guerrieri and Shimer (2014, 2018), Chang (2017), Li (2019), Williams (2019), and Kargar et al. (2020).

2 The Model

Agents, assets, and preferences. Time is continuous and runs forever. The economy is populated by a unit measure of infinitely-lived, risk-neutral investors who discount the future at rate $r > 0$. There is one durable asset in fixed supply $A \in (0, 1)$ and one perishable numéraire good with marginal utility normalized to one. Investors can hold either zero or one unit of the asset, which is assumed to be indivisible. We refer to investors who have the asset as "owners" and investors who do not as "non-owners". The instantaneous utility flow investors receive from holding the asset is δ_i, where $i \in \{1, 2, ..., I\} \equiv \mathcal{I}$ indexes an investor's type with $1 < I < \infty$, and $\delta_i > \delta_j$ for $i > j$. We postulate discrete types, as the majority of search-theoretic models of OTC markets, because this guarantees that all intermediation chains in our model will have finite length. Types change over time: each investor receives i.i.d preference shocks according to a Poisson process with intensity γ. Conditional on receiving a preference shock, the investor draws a new type j from some discrete cumulative distribution function $F(\delta_j)$ with support $\mathcal{I}$. We denote the probability mass function of that distribution by $f(\delta_j)$. The measures of owners and non-owners type i are $o(\delta_i)$ and $n(\delta_i)$ respectively.

Matching and trade. Investors interact in a fully decentralized market: all trades take place in bilateral meetings. The market features complete price transparency: each owner posts (and commits to) a publicly available price at which she is willing to sell the asset to a non-owner.[9] Non-owners observe all available prices and direct their orders to at most one of the available prices. In this regard we follow Moen (1997), Acemoglu and Shimer (1999), and Eeckhout and Kircher (2010) and model the market using a *competitive search* protocol. Moreover, we assume the types of all agents are publicly observable. We refer to the collection of all owners posting the same price and all non-owners willing to buy the asset

[9] We should emphasize that focusing on price posting instead of allowing owners to post more complicated contracts is without any loss of generality, given the matching protocol we use.

at this price as a "submarket".[10] Owners (non-owners) who participate in some submarket we call sellers (buyers) on that submarket. Owners (non-owners) who do not participate in any submarket, we refer to as idle.

In practice, it takes time to execute trades between investors because of frictions. For example, there are informational asymmetries (it takes time to verify collateral), institutional constraints (on the volume of trade), and technological limitations on the speed of execution. We follow Lagos and Rocheteau (2009) and Lester et al. (2015) and capture these delays through the means of a matching function. Consequently, the speed with which counterparties trade is endogenous and depends on the queue length (the ratio of buyers and sellers) on each submarket.[11] Formally, suppose there is a measure $o(\delta_i, p)$ of owners type i posting a particular price p and a measure $n(\delta_j, p)$ of non-owners type j interested in acquiring a unit of the asset at this price. Then, at each instant, the flow of trades executed is given by $m\left(\sum_i o(\delta_i, p), \sum_j n(\delta_j, p)\right)$, a function which has constant returns-to-scale, is strictly increasing, strictly concave and twice continuously differentiable with respect to its two arguments. As a consequence, the waiting time for an owner to sell her unit of the asset is an exponentially distributed random variable with parameter $\lambda(q(p)) \equiv \frac{m\left(\sum_i o(\delta_i,p), \sum_j n(\delta_j,p)\right)}{\sum_i o(\delta_i,p)} = m\left(1, q(p)\right)$, where $q(p) \equiv \frac{\sum_j n(\delta_j,p)}{\sum_i o(\delta_i,p)}$ is the queue length on the submarket. Symmetrically, the waiting time for a non-owner is exponentially distributed with intensity $\frac{\lambda(q(p))}{q(p)}$. The assumptions on the matching function imply that $\lambda(\cdot)$ is continuous, strictly increasing, and strictly concave.

[10] Moen (1997) and Mortensen and Wright (2002) assume that third-party market makers set up submarkets and promise a price and expected waiting time for transactions to any agents that show up. We follow the formalization of Acemoglu and Shimer (1999) and Eeckhout and Kircher (2010) in which sellers post terms of trade and buyers self-select to different submarkets. These are all equivalent interpretations of the competitive search protocol; see Wright et al. (2019).

[11] To be precise, the queue length equals the ratio of non-owners over owners along the equilibrium path. Out of the equilibrium path, the queue length will be determined by a condition on investors' beliefs; we explain the condition precisely later in Section 3.1.

3 Equilibrium

3.1 Value Functions and Equilibrium Definition

Owners. Let $V_1(\delta_i, p)$ denote the expected lifetime payoff of an owner of type i who posts a price p, expecting a queue length $q(p)$:

$$rV_1(\delta_i, p) = \delta_i + \gamma \sum_j \left[V_1^*(\delta_j) - V_1(\delta_i, p)\right] f(\delta_j) + \lambda(q(p)) \left[p - V_1(\delta_i, p) + V_0^*(\delta_i)\right]. \quad (1)$$

Intuitively, the owner enjoys a utility flow δ_i from holding the asset until she receives a preference shock or sells the asset. At a rate γ, the owner draws a new preference type j with probability $f(\delta_j)$. In this event, she obtains the maximum attainable utility of being an owner of type j, denoted by $V_1^*(\delta_j) \equiv \max_{p\in\mathcal{P}} \{V_1(\delta_j, p)\}$, where $\mathcal{P}$ is the set of prices posted in equilibrium, and loses her current expected payoff, $V_1(\delta_i, p)$. At a rate $\lambda(q(p))$, she meets a non-owner who purchases the asset at price p. In this event, the owner receives the price and the maximum attainable utility of being a non-owner of type i, denoted by $V_0^*(\delta_i) \equiv \max_{p\in\mathcal{P}} \{V_0(\delta_i, p)\}$, but loses her current expected payoff, $V_1(\delta_i, p)$.

Non-owners. Next, consider a non-owner type i who participates in submarket p:[12]

$$rV_0(\delta_i, p) = \gamma \sum_j \left[V_0^*(\delta_j) - V_0(\delta_i, p)\right] f(\delta_j) + \frac{\lambda(q(p))}{q(p)} \left[V_1^*(\delta_i) - p - V_0(\delta_i, p)\right]. \quad (2)$$

Intuitively, a non-owner with type i does not enjoy a positive utility flow since she does not own the asset. Two distinct events can affect her value function: i) the preference shock and ii) meeting an owner. At rate γ, the non-owner draws a new preference type and this type is j with probability $f(\delta_j)$. In this event, she obtains the maximum attainable utility of being a

[12] If she finds all posted prices too high, she chooses to not participate in any submarket, i.e. $p = \emptyset$. This notation allows us to keep the exposition concise and denote the value function of non-owners who are active buyers in some submarket and non-owners who are idle by $V_0(\delta_i, p)$. If the non-owner chooses the option of $p = \emptyset$, her matching rate with an owner is 0, i.e. $\lambda(q(\emptyset))/q(\emptyset) = 0$.

non-owner of type j, $V_0^*(\delta_j)$, but loses her current payoff, $V_0(\delta_i, p)$. At rate $\lambda(q(p))/q(p)$, she meets an owner and purchases the asset. In that case, she receives the maximum attainable utility of being an owner of type i, $V_1^*(\delta_i)$, but loses her current expected payoff.

Out-of-equilibrium beliefs. So far, the value functions are only determined along the equilibrium path, since the queue length, $q(p)$, is well defined only for prices $p \in \mathcal{P}$. We follow Eeckhout and Kircher (2010) and Jerez (2014) by imposing restrictions on beliefs in the spirit of subgame perfection: owners expect a positive queue length when posting a given price only if there is some non-owner who is willing to buy the asset at that price. Moreover, owners expect the largest possible queue length for which they can find such a non-owner type, i.e. they expect non-owners to queue in the submarket until it is no longer profitable to do so. Formally, the queue length satisfies the following condition:

$$q(p) = \sup \left\{ q' \in \mathbb{R}_+ : \exists \delta; \quad V_0(\delta, p, q') \geq \max_{p' \in \mathcal{P}} V_0(\delta, p') \right\}, \tag{3}$$

or $q(p) = 0$, if the set is empty.[13] This belief restriction defines queue lengths and value functions on the entire domain of prices, not only along the equilibrium path.

Laws of motion. The masses of agent types evolve due to preference shocks and trading:

$$\dot{o}(\delta_i) = \gamma f(\delta_i) \sum_{j \neq i} o(\delta_j) - \gamma o(\delta_i) \sum_{j \neq i} f(\delta_j) - \sum_{p \in \mathcal{P}} o(\delta_i, p) \lambda(q(p)) + \sum_{p \in \mathcal{P}} n(\delta_i, p) \frac{\lambda(q(p))}{q(p)}, \tag{4}$$

$$\dot{n}(\delta_i) = \gamma f(\delta_i) \sum_{j \neq i} n(\delta_j) - \gamma n(\delta_i) \sum_{j \neq i} f(\delta_j) + \sum_{p \in \mathcal{P}} o(\delta_i, p) \lambda(q(p)) - \sum_{p \in \mathcal{P}} n(\delta_i, p) \frac{\lambda(q(p))}{q(p)}. \tag{5}$$

The first term represents the net flow of agents into $o(\delta_i)$ due to preference shocks: at rate γ owners are hit with a preference shock and conditional on that, each agent draws type i with probability $f(\delta_i)$. The second term captures the net flow out due to preference shocks: each owner of type i receives a preference shock at rate γ and, conditional on that event,

[13] The notation $V_0(\delta, p, q')$ emphasizes the value function depends on the queue length q'.

her type changes with probability $\sum_{j\neq i} f(\delta_j)$. In each submarket p there are $o(\delta_i, p)$ owners of type i selling the asset. Each of them executes a trade at rate $\lambda(q(p))$. Once the trade is executed the owner transfers the asset to the non-owner and she becomes a non-owner of type i. Thus, the third term in equation (4) captures the flow of agents out of $o(\delta_i)$ due to trade. Analogously, the last term captures the flow into owners of type i due to trade. The law of motion for non-owners is given by (5) and the intuition behind it is analogous.

Equilibrium definition. A steady state equilibrium is a set of value functions $V_1^*(\delta_i)$ and $V_0^*(\delta_i)$, with $i \in \mathcal{I}$, a set of prices $\mathcal{P}$, a set of masses $o(\delta_i, p)$ and $n(\delta_i, p)$, with $p \in \mathcal{P}$ and $i \in \mathcal{I}$, and a queue length function $q(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$, such that owners choose prices that maximize their values defined by (1), non-owners choose submarkets that maximize their values defined by (2), the queue length function satisfies condition (3), equations (4) and (5) hold with $\dot{o}(\delta_i) = \dot{n}(\delta_i) = 0$, for all $i \in \mathcal{I}$, the accounting identities $\sum_{p\in\mathcal{P}} o(\delta_i, p) = o(\delta_i)$ and $\sum_{p\in\mathcal{P}} n(\delta_i, p) = n(\delta_i)$ hold for all $i \in \mathcal{I}$ and the resource constraints $\sum_{i\in\mathcal{I}} o(\delta_i) = A$ and $\sum_{i\in\mathcal{I}} n(\delta_i) = 1 - A$ hold.

Value functions along the equilibrium path. The equilibrium payoff of each investor is determined by her own prices and submarket choices and by the *distribution* of price and submarket choices in the economy, which in turn has to be consistent with the optimal choices of individual investors. In equilibrium, an individual owner of type i takes the distributions of investor masses $o(\delta_i, p)$ and $n(\delta_i, p)$ as given and, according to the equilibrium definition, her pricing decision solves $\max_p V_1(\delta_i, p)$ (taking into account that her choice of price affects the expected queue length, q). The owner can post a price that attracts a zero ($q(p) = 0$) or positive ($q(p) > 0$) measure of non-owners. The response of non-owners to this seller's pricing decision is captured by (3), which holds by assumption outside the set of equilibrium prices $\mathcal{P}$ and, by the equilibrium definition, inside $\mathcal{P}$. Thus, the problem of an owner is

$$\max_{q,p} \left\{\lambda(q)\left[p - V_1(\delta_i, p) + V_0^*(\delta_i)\right] : q = \sup\left\{q' \in \mathbb{R}_+ : \exists\delta;\ V_0(\delta, p) \geq V_0^*(\delta)\right\}\right\}. \tag{6}$$

It is easy to establish that, by continuity of $V_0(\delta_i, p)$ in (6), any optimally set price leaves the non-owners who queue up in this submarket their market utility $V_0^*(\delta)$. As a result, the owner's i problem reduces to

$$\max_{q,p,j} \{\lambda(q) [p - V_1(\delta_i, p) + V_0^*(\delta_i)] : V_0(\delta_j, p) = V_0^*(\delta_j)\}. \tag{7}$$

Variants of equation (7) are at the core of most competitive search models. Intuitively, in equilibrium, owners optimally choose prices and non-owner type(s) they wish to attract, such that non-owners receive exactly the utility they would get if they were to participate in a different submarket. $V_0^*(\delta)$ is an endogenous object and is taken as given by individual investors. The existing literature has used this strategy of characterizing the equilibrium (often referred to as *market utility approach*) extensively; see Wright et al. (2019).

The Bellman equations along the equilibrium path read:

$$(r+\gamma)V_1^*(\delta_i) = \delta_i + \gamma V_1^* + \lambda(q^*(\delta_i)) [p^*(\delta_i) - \Delta V^*(\delta_i)], \tag{8}$$

$$(r+\gamma)V_0^*(\delta_i) = \gamma V_0^* + \frac{\lambda(q^{**}(\delta_i))}{q^{**}(\delta_i)} [\Delta V^*(\delta_i) - p^{**}(\delta_i)], \tag{9}$$

where $V_1^* = \sum_j V_1^*(\delta_j) f(\delta_j)$ and $V_0^* = \sum_j V_0^*(\delta_j) f(\delta_j)$ are the average maximum utilities across investor types, $p^*(\delta_i)$ is an optimally chosen price posted by owner i that attracts queue length $q^*(\delta_i)$ in equilibrium, $p^{**}(\delta_i)$ and $q^{**}(\delta_i)$ are the equilibrium price and queue length in a submarket that non-owner i has optimally chosen to participate and $\Delta V^*(\delta_i) = V_1^*(\delta_i) - V_0^*(\delta_i)$ is the *reservation value* of investor i (that is, the expected utility of owing minus the expected utility of not owing the asset) along the equilibrium path.[14]

Reservation values along the equilibrium path. We can use the expressions for the

[14] Of course, there may be more than one price and more than one submarket that maximize an owner's and a non-owner's lifetime utility respectively.

Bellman equations, (8) and (9), to express the reservation value of investor type i as

$$(r+\gamma)\Delta V^*(\delta_i) = \delta_i + \gamma \Delta V^* + \lambda(q^*(\delta_i))\left[p^*(\delta_i) - \Delta V^*(\delta_i)\right] - \frac{\lambda(q^{**}(\delta_i))}{q^{**}(\delta_i)}\left[\Delta V^*(\delta_i) - p^{**}(\delta_i)\right]. \tag{10}$$

This expression highlights the dual role of investors (as buyers and sellers) in the market. The first two terms capture the utility flow of owning the asset and the expected capital gain (loss) from switching types. The last two terms capture the search options of the investor when she is an owner and a non-owner. When she does not have the asset, the investor can queue on the submarket $p^{**}(\delta_i)$ and, if matched, she purchases the asset. She then receives her reservation value and is now an owner who can exercise her option to post the asset for sale. If the optimally chosen $p^*(\delta_i)$ attracts a positive queue length, once the investor makes contact with a buyer, she receives the price but loses her reservation value. Thus, she is now a non-owner and can again exercise her search option of buying asset.

Following Hugonnier et al. (2021), we call this behavior *intermediation*. It is important to stress that intermediation in our model is a choice. The sellers on submarket $p^{**}(\delta_i)$ choose to post terms of trade such that they trade with an investor type i who then sells the asset to buyers on submarket $p^*(\delta_i)$, even though sellers on market $p^{**}(\delta_i)$ can post terms of trade that attract the non-owners on market $p^*(\delta_i)$. Similarly, non-owners on market $p^*(\delta_i)$ choose to buy the asset from an investor type i rather than to queue directly on the market $p^{**}(\delta_i)$. Thus, any intermediation that takes place in our economy is an outcome of investors' optimal choices regarding trading counterparties. In contrast, in the random search framework intermediation is a mechanical consequence of the matching technology.[15]

[15] For example, in Hugonnier et al. (2021) investors with moderate valuations intermediate trades for investors with extreme valuations, but this is an implication of random meetings taking place with an exogenous intensity. Moderate types meet relatively frequently with both non-owners with higher types than their own and owners with lower types than their own and, as a result, intermediate a large fraction of trades. In our model, low types *choose* to search for and sell to moderate types and moderate types *choose* to search for and sell to high types. Hence, intermediation is not a by-product of random exogenous meetings and asset misallocation across investor types, but rather arises naturally as investors direct their search towards different prices. Of course, in the random search model investors can choose whether or not

Trade surplus along the equilibrium path. Whenever investors trade, this generates a surplus. We denote it by $S^*(\delta_i, \delta_j)$, where i, j are the types of the seller and buyer respectively and the star superscript represents the value along the equilibrium path. If these two agents trade, the buyer gains her reservation value, but transfers the price to the seller who in turn loses her reservation value. Hence, $S^*(\delta_i, \delta_j) = \Delta V^*(\delta_j) - \Delta V^*(\delta_i)$. Using (10) the surplus can be rewritten as

$$\begin{aligned} S^*(\delta_i, \delta_j) = \frac{\delta_j - \delta_i}{r+\gamma} + \frac{\lambda(q^*(\delta_j))}{r+\gamma}\left[p^*(\delta_j) - \Delta V^*(\delta_j)\right] + \frac{\lambda(q^{**}(\delta_i))}{q^{**}(\delta_i)(r+\gamma)}\left[\Delta V^*(\delta_i) - p^{**}(\delta_i)\right] \\ - \frac{\lambda(q^*(\delta_i))}{r+\gamma}\left[p^*(\delta_i) - \Delta V^*(\delta_i)\right] - \frac{\lambda(q^*(\delta_i))}{q^*(\delta_i)(r+\gamma)}\left[\Delta V^*(\delta_j) - p^*(\delta_i)\right]. \quad (11) \end{aligned}$$

The above expression highlights how the option of acting as an intermediary affects the trade surplus. The first term of (11) captures the discounted difference in the utility flows of the two investors. We call this the *fundamental surplus*. The next two terms capture the potential benefit for the two investors if they do act as intermediaries. When the buyer receives the asset she not only gains her flow value but also the option to post the asset for sale on a different submarket $p^*(\delta_j)$. At the same time the owner i gains the option to search for the asset her self and reacquire it at the price $p^{**}(\delta_i)$. Both of these options have positive payoffs and so they increase the surplus of the match. If the buyer can sell the asset at favourable terms quickly or if the seller can reacquire the asset quickly at a low price, this increases the surplus of their match. The last two terms capture the loss of the investors' outside option once they trade.

trade takes place once they make contact. However, it is always optimal to trade, provided the surplus is positive. Thus, intermediation properties in that model are purely a mechanical consequence of exogenous matching intensities.

3.2 Equilibrium Characterization

Next, we proceed to characterize the equilibrium (we defer equilibrium existence to Appendix A). The first set of equilibrium properties establishes that investors' value functions and reservation values are increasing in investor types, a result which is standard in the literature.

Lemma 1. *For any $i > j$, the following inequalities hold: $\Delta V^*(\delta_i) > \Delta V^*(\delta_j)$; $V_0^*(\delta_i) \geq V_0^*(\delta_j)$; $V_1^*(\delta_i) > V_1^*(\delta_j)$. Furthermore, if some non-owners of type j are active in some submarket, then $V_0^*(\delta_i) > V_0^*(\delta_j)$.*

Next, we can use equations (8) and (9), along with the constraint $V_0(\delta_j, p) = V_0^*(\delta_j)$, to express the owner's problem (7) more compactly:

$$\max_{q,j} \left\{ \lambda(q) S(\delta_i, \delta_j) - q \left[(r + \gamma) V_0^*(\delta_j) - \gamma V_0^* \right] \right\}, \tag{12}$$

where $S(\delta_i, \delta_j) = \Delta V^*(\delta_j) - \Delta V(\delta_i, p)$. The first order condition with respect to the queue length results in an expression for the price of trade between i and j, familiar from the competitive search literature:

$$p(\delta_i, \delta_j) = \eta(q) \Delta V^*(\delta_j) + [1 - \eta(q)] \Delta V(\delta_i, p), \tag{13}$$

where $\eta(q) \equiv \frac{o m_o}{m}$ is the elasticity of the matching function with respect to the total measure of owners in this submarket. In equilibrium, the price of the asset is a weighted average of the two equilibrium reservation values and thus captures investors' option to act as intermediaries. The reservation value of the buyer is higher if she can sell the asset quickly at a favorable price, once she has acquired it. This increases the surplus of the match and consequently the price. As a result some of the potential benefits the buyer receives by acting as an intermediary are passed to the seller at the time of the trade. Similarly, if the seller has to reacquire the asset, once she has sold it, at relatively high prices or slow speeds, the expected "costs" or restocking her inventory are high. This decreases the match surplus and the price. Thus, some of the expected costs of reacquiring the asset are passed to the buyer.

It is instructive to compare (13) with equation (3) of Hugonnier et al. (2021): they are exactly the same, except the weights on $\Delta V^*(\delta_j)$ and $\Delta V(\delta_i, p)$ are the bargaining powers of owners and non-owners respectively instead of the matching function elasticity.[16] In the model of Hugonnier et al. (2021), the price of trade is the result of Nash bargaining between owners and non-owners with different bargaining powers. In both models, the price is a weighted sum of the reservation values of investors who trade, but the weights are different: with Nash bargaining the weights are exogenous and equal to investors' bargaining powers, while with price posting the weights are endogenous and depend on the queue length of the submarket in which the trade takes place. To better understand the implications of this result, we use (13) to express the gains from a trade between owner type i and non-owner type j as functions of the match surplus: $p(\delta_i, \delta_j) - \Delta V(\delta_i, p) = \eta(q)S(\delta_i, \delta_j)$ and $\Delta V^*(\delta_j) - p(\delta_i, \delta_j) = [1 - \eta(q)]S(\delta_i, \delta_j)$. These expressions correspond to the Hosios (1990) condition for efficiency in markets with search frictions. However, as discussed by Shi (2001) and Eeckhout and Kircher (2010), in environments with two-sided heterogeneity, equation (13) is necessary but not sufficient for the equilibrium to be efficient.[17] We show that the equilibrium in our model is actually efficient in Appendix A.

Our results on the price and surplus along the equilibrium path allow us to show that in equilibrium agents will endogenously segment into different submarkets. That is, every submarket features a unique pair of seller and buyer types, a typical implication of com-

[16] All search and matching models of OTC trading with ex post bargaining feature a version of this equation; for example, it is equation (11) in Duffie et al. (2005).

[17] Condition (13) dictates how a given pair (δ_i, δ) should split the match surplus to achieve efficiency. It does not specify *which pairs* (δ_i, δ) should trade in equilibrium. That is, with two-sided heterogeneity the question of which pairs are formed in equilibrium is crucial for efficiency; this question is trivial in models with homogeneous agents or with one-sided heterogeneity.

petitive search. Intuitively, price competition among sellers drives the price to the weighted sum of reservation utilities of owners and non-owners that trade in the submarket. Hence, if there are two buyer types in a submarket, one of them does not receive their market utility (since these are strictly increasing in investors' types). This contradicts optimal behavior of non-owners. Symmetrically, when two sellers of different types post the same price, there is a profitable deviation for one of them.

Proposition 1. *Every submarket in equilibrium features only one type of owner and one type of non-owner. Furthermore, if owner i and non-owner j trade in equilibrium, they trade in only one submarket.*

As a consequence, we can index submarkets by (δ_i, δ_j), where i is the owner type and j is the non-owner type in this submarket. Accordingly, we write $q(\delta_i, \delta_j)$ and $p(\delta_i, \delta_j)$ for submarket (δ_i, δ_j). We use the star superscript to denote variables along the equilibrium path: $q^*(\delta_i, \delta_j)$ and $p^*(\delta_i, \delta_j)$. Even though each submarket is populated by a unique pair of seller-buyer types, agents from a given seller (buyer type) may participate on many different submarkets, i.e. it is possible for both submarkets (δ_i, δ_j) and (δ_i, δ_k) to open.

3.3 Intermediation

We are now ready to turn our attention to the main result of the paper: in our environment, intermediation arises naturally and is ubiquitous. Following our definition, an agent acts as an intermediary if she buys the asset when she does not have it and sells it when she does. This essentially translates to participating in submarkets as both active buyer and active seller. Thus, a natural next step is to examine which submarkets open in equilibrium.

Proposition 2. *In equilibrium, no market (δ_i, δ_j) with $i > j$ opens. Furthermore, if submarket (δ_i, δ_j) with $i < j$ opens, then markets (δ_i, δ_i) and (δ_j, δ_j) do not open.*

As expected, pairs with negative fundamental surplus ($i > j$) do not trade in equilibrium. Moreover, the only occasion in which zero-fundamental-surplus pairs of investors ($i = j$) occurs is when this type does not participate in any submarket with positive surplus. If there is no opportunity for trades with positive surplus, investors are indifferent between non-participation and trading with investors with the same valuation. In such a situation, we assume that when owners are indifferent between being idle and participating in some submarket, they post a price which attracts a positive queue length. Given the off-equilibrium beliefs, this implies that if there are both owners and non-owners of some type i that do not participate in any market with a positive surplus they participate in (δ_i, δ_i). Furthermore, the next corollary follows immediately.

Corollary 1. *If some owners (non-owners) of a given type i are sellers (buyers) on some submarket, then all owners (non-owners) of that type are sellers (buyers) on some submarket.*

Proof. The proof is immediate from Proposition 2. ■

Of course in equilibrium not all submarkets with positive surplus necessarily open. In general, answering the question of which submarkets open depends on parameter values and on the functional form of the matching function. In this Section, we retain our general framework and focus on participation, i.e. which owners are going to be active sellers and which non-owners would be active buyers. This approach allows us to explore the existence of intermediation in our model without making restrictive parametric assumptions. In Sections 4 and 5, we provide more detailed results for specific commonly-used matching functions.

The potential benefit to investors from participating in a submarket is affected by their reservation value. When a non-owner buys the asset, she pays the price but receives her reservation value. Since equilibrium reservation values are strictly increasing in the investor's type, it follows that higher type non-owners have a stronger inventive to participate in some submarket as compared to lower types. Similarly, when an owner sells the asset she loses her reservation value. As a consequence, the higher the type of the owner, the lower her incentive to participate in some submarket. The next proposition formalizes this intuition.

Proposition 3. *In equilibrium, an owner type i is an active seller if and only if $i \leq s$ and a non-owner type j is an active buyer if and only if $j \geq b$, where s and b are some types between* 1 *and* I.

In equilibrium, all non-owners of high enough valuation are active buyers and all owners of low enough valuation are active sellers. Intermediation would arise if $s \geq b$. In that situation, there would be some types of investors who are actively selling the asset when they have it, and actively trying to acquire the asset when they do not have it. Whether or not this is true, however, depends on the primitives of the model, and in particular, on the properties of the matching function.

Proposition 4 provides a clear characterization of the case of extensive intermediation. If the matching function satisfies an asymptotic condition, reminiscent of the Inada conditions commonly used in Macroeconomics, then almost every investor type (other than the highest and lowest types) acts as an intermediary. These Inada-type conditions are satisfied by many matching technologies commonly used in the search and matching literature, such as the Cobb-Douglas matching function. Then proposition 4 implies that intermediation is a natural outcome in asset models with matching frictions and does not rest on either random meetings or ex post bargaining. Matching frictions are enough to generate intermediation, even when prices are posted before trade takes place. In Sections 4 and 5, we also show the configuration of the market structure and intermediation in the competitive search equilibrium for specific matching technologies, some of which do not satisfy the Inada-type conditions.

Proposition 4. *If the matching function satisfies the Inada-type conditions* $\lim_{q\to\infty} \lambda(q) = \lim_{q\to 0} \lambda(q)/q = \infty$, *then* $b = 2$ *and* $s = I - 1$.

The proposition above delivers our central result: intermediation is a natural outcome in

OTC models with matching frictions, even when agents can choose trading counterparties and terms of trade. To understand the intuition, let us first look at it through the planner's point of view. Consider three types of agents: low, mid, and high (l, m, h), with $\delta_l < \delta_m < \delta_h$. Among these three types, the largest welfare gains are generated when an owner type l transfers the asset to a non-owner type h. The planner can instruct these agents to trade directly, in which case it will take on average $1/\lambda(q(l,h))$ time for the owner to offload the asset. Alternatively, she can instruct m to act as an intermediary: in that case the total time it takes for the asset to reach type h is $1/\lambda(q(l,m)) + 1/\lambda(q(m,h))$. Thus, intermediation comes at a price: the asset will generally take longer to reach the high type agent. At the same time, there is a benefit to intermediation: for part of this extra time the asset is held by type m who values it more than type l. As a result, the planner will find it optimal for intermediation to take place if $\delta_m - \delta_l$ (the benefit) is large enough, and the trade delay (the cost) small enough. When the matching function satisfies the Inada-type conditions, $\lim_{q\to\infty} \lambda(q) = \infty$, the planner can achieve almost infinite execution speeds if she assigns a large enough mass of non-owners and small enough mass of owners on submarket (l,m). She can then instruct the agents of type m who receive the asset to transfer it to agents of type h on a submarket with a queue length $q(m,h) = q(l,h)$. Since l transfers the asset to m almost instantenously, this means that there is no trading delay due to intermediation, i.e. $1/\lambda(q(l,h)) \approx 1/\lambda(q(l,m)) + 1/\lambda(q(m,h))$. However, there are strictly positive welfare gains because, while waiting to transfer, the asset owners of type m enjoy a flow benefit δ_m which is strictly larger than the flow benefit δ_l the low type owner would have enjoyed.[18]

Price posting guarantees that the above argument is implemented in the competitive

[18] This intuition, based on efficiency, is analogous to the intuition behind why negative assortative matching arises in a static assignment problem; see Eeckhout and Kircher (2010). It is more efficient to have *both* low and high types trade fast which, due to matching frictions, implies they must trade with medium valuation agents, as opposed to trading directly with each other. We thank an anonymous referee for pointing this out. This analogy becomes even more transparent in the context of our Lemma 3 which implies that positive assortative matching lowers aggregate surplus.

search equilibrium, because prices allow investors to share the gains from intermediation. For intermediation to occur in equilibrium, it must be the case that the owner type l and non-owner type h both get favorable terms of trade when they trade with type m, even though the surplus generated in these trades is smaller than the surplus generated in trades between l and h. Thus, it must be the case that l (h) extracts a large fraction of the surplus when she trades with non-owner (owner) type m, or that she can execute a trade relatively fast. On the other hand, investors type m always have an incentive to act as intermediaries. If they do not intermediate, they are idle; if they are active, there is the chance to execute a positive surplus trade and extract some of that surplus. Thus, intermediation arises when types m can provide favorable enough terms of trade and fast order execution speeds. When the matching function satisfies the Inada-type conditions, m can provide almost instantaneous trades to both l and h and at the same time, through prices, transfer enough of the intermediation gains to incentivize these trades.

Proposition 4 characterizes the case of *extensive* intermediation: the Inada-type conditions guarantee that all agent types $2 \leq m \leq I-1$ act as intermediaries, regardless of how small $\delta_m - \delta_l$ might be. Intermediation is ubiquitous because there are no costs to it. On the other hand, if the matching function does not satisfy the Inada-type conditions and the expected waiting times for an order execution are bounded away from zero, there will be positive opportunity costs in intermediation. As a result, the extent of intermediation may be limited: only intermediation links that provide large enough gains will be formed.

The degree of market intermediation may also be limited if there are flow costs from participating in a submarket. In that case, the costs of participating for the intermediary may outweigh the extra surplus from intermediation. Similar to the previous case, only intermediated trades that generate enough large additional surplus will be formed. Finally, intermediation may not arise in the frictionless limit of the model. As trading delays go to zero, so do the benefits from intermediation.

When the matching function does not satisfy the Inada-type conditions or when there are participation costs, intermediation is still likely to exist albeit to a lesser extent. In these cases, intermediation is a costly activity and only agents that can create a large enough improvement in the surplus will act as intermediaries. For example, in Section 5, we show that even when the matching function does not satisfy the Inada-type conditions, dealers intermediate more than 70% of trades for most parameter values. Even though intermediation may be less ubiquitous under certain matching functions, we can show that for any matching function all investors types will participate in the market in some capacity.

Lemma 2. *In equilibrium, for any type i, either all owners are sellers on some submarket, or all non-owners are buyers on some submarket, or both.*

So far, we have examined the extent of participation in the market and its relation to intermediation. Proposition 4 also sheds light on the intermediation patterns observed in equilibrium. These patterns essentially describe how agents of different types *sort* into submarkets. A vast literature has examined bilateral trade between heterogeneous agents in the context of sorting problems: see, for example, Shi (2001), Eeckhout and Kircher (2010) and Jerez (2014); a recent survey is provided in Chade et al. (2017). In this context, a sorting function takes as inputs non-owner types and assigns them to owner types for trading. The pairwise matching of investor types induced by a sorting function is called an assignment. Following Eeckhout and Kircher (2010), *assortative matching* includes only strictly monotone assignments. That is, an assignment is called positive (negative) assortative matching, or simply PAM (NAM), if the sorting function that assigns a non-owner type to an owner type in equilibrium is strictly decreasing (increasing).[19]

[19] In the sorting literature, an assignment is referred to as positive (negative) if high valuation buyers match with high (low) valuation sellers. In this context, the ordering of buyer and seller types is such that the match surplus is increasing in both types. On the contrary, in our model the fundamental surplus is decreasing in the seller's valuation. As a result, for our purposes, what the literature has traditionally referred to as a "high type seller" corresponds to an owner with low valuation of the asset.

Lemma 3 summarizes the implications of our intermediation results for the possible equilibrium sorting patterns. Specifically, our result that almost all agent types act as intermediaries puts very tight constraints on the sorting patterns that can be observed in equilibrium. Firstly, PAM cannot be an equilibrium, since it essentially precludes intermediation. Importantly, given that the equilibrium is constrained efficient, this result implies that PAM lowers aggregate welfare. Second, NAM can be an equilibrium outcome, but in this case the asset must move from the lowest to the highest valuation investor via a single chain that includes every single investor type.

Lemma 3. *Suppose the matching function satisfies* $\lim_{q\to\infty}\lambda(q) = \lim_{q\to 0}\lambda(q)/q = \infty$ *and that* $I > 2$*. Then, PAM cannot be an equilibrium. Moreover, the only NAM equilibrium is a single chain which starts at owner type* 1 *and ends at non-owner type* I*. That is, all possible submarkets of the type* (δ_i, δ_{i+1}) *open and no other submarket opens.*

4 Analytical Examples

The main result of the paper, namely that intermediation always occurs in equilibrium, relies on the properties of the matching function. If the matching function does not satisfy the Inada-type conditions, an intermediary may not be able to move the asset fast enough from one customer to another for intermediation to be efficient. As a result, the extend of intermediation may be limited, depending on the parameter values of the model and the exact functional form of the matching technology. In this section, we further explore how equilibrium intermediation properties depend on the matching technology and model parameters with three illustrative examples for which we can obtain closed-form solutions. In all three examples we restrict the number of investor types to three, set the probability of type switching to $f(\delta_1) = f(\delta_2) = f(\delta_3) = 1/3$, and normalize the investor valuations to $\delta_1 = 0, \delta_2 = \delta, \delta_3 = 1$. In the first two examples we consider a matching technology linear in

Matching / Asset	Leontief	Linear	
	$m(o,n) = \mu \min\{o,n\}$	$m(o,n) = \mu o$	$m(o,n) = \mu n$
$A \leq 1/3$	$(1,3) + (2,3) > (1,2) + (2,3)$	$(1,3) + (2,3)$	$(1,2) + (1,3)$
$1/3 < A < 1/2$	All $= (1,3) + (2,3) \geq (1,2) + (2,3)$		
$A = 1/2$	All $= (1,2) + (2,3)$		
$1/2 < A < 2/3$	All $= (1,2) + (1,3) \geq (1,2) + (2,3)$		
$2/3 \leq A$	$(1,2) + (1,3) > (1,2) + (2,3)$		

Table 1: Market Structure and Welfare for Linear and Leontief Matching Functions

the mass of (i) sellers and (ii) buyers, i.e. (i) $m(o,n) = \mu o$ and (ii) $m(o,n) = \mu n$, where μ is a matching efficiency parameter. In the last example we turn our attention to a Leontief matching technology: $m(o,n) = \mu \min\{o,n\}$.

The details of how we solve for the equilibrium are contained in Appendix C, but we provide a brief overview here. We take a two-step approach. First, we analyze each possible combination of submarkets that may open, i.e. the market structure, one at a time and derive the set of parameter values consistent with that particular market structure. Second, we rank the feasible market structures in terms of welfare for each set of parameter values identified in the first step. Given that the equilibrium is constrained efficient, the highest ranked market structure will be the equilibrium.[20]

Table 1 summarizes our findings. Although the speed of trade and prices depend on all model parameters, in the case of a linear matching function the market structure is

[20] The Leontief matching function is not differentiable at $q = 1$, so our Proposition 6 from the previous section is not directly applicable. As a result, we have to explicitly verify that the highest ranked market structure will indeed be an equilibrium. We confirm this in the Appendix. In addition, neither the Leontief nor the linear matching functions are strictly increasing in the masses of both buyers and sellers. Because of this the equilibrium may feature agents of different types trading at the same prices, i.e. submarkets may have more than one type of buyer or seller. To make the comparison with our analysis in the previous sections transparent and present our results on the pattern of trade in the clearest way possible, in this section we assume that sellers post terms of trade directed towards a specific buyer type and buyers direct their orders towards a specific seller type. This assumption allows us to have an equilibrium market structure where each submarket features only one type of seller and one type of buyer, just as in the previous section. Moreover, this assumption has no effect on equilibrium prices or speeds at which agents trade.

independent of parameter values. When the matching function is linear in the mass of sellers the only possible market structure is submarkets $(1, 3)$ and $(2, 3)$ open, while submarket $(1, 2)$ does not. When it is linear in the mass of buyers, only submarkets $(1, 2)$ and $(1, 3)$ open. When the matching function is Leontief, the market structure depends only on the level of the asset supply. Under extreme asset levels, either very large or very small, the equilibrium features no intermediation. For intermediate quantities of the asset supply, intermediation emerges and all markets open. However, except for the knife-edge case of $A = 1/2$, intermediation is not strictly welfare improving. Our results then imply that (i) a necessary condition for intermediation to emerge is to have congestion on both sides of the market; and (ii) the welfare-improving effects of intermediation may be limited.

To understand the intuition behind our results, let us first focus on the logic behind why intermediation is never efficient when only one side of the market contributes to congestion, i.e. the matching function is linear. Suppose the matching function is linear in the mass of seller (the logic for buyers is similar). In that case, the seller's matching rate, $\lambda(q) = \mu$, is constant, but the buyer's matching rate is decreasing in q. Then, the efficient allocation is to have the queue lengths as low as possible so that the asset reaches high valuation buyers as quickly as possible. This is achieved when all owners sell to type 3 non-owners and only submarkets $(1, 3)$ and $(2, 3)$ open. If submarket $(1, 2)$ were to open, this would result in an increase in $q(\delta_1, \delta_3)$ and, as a result, lower order execution rates for type 3 buyers. Thus, the asset takes longer to reach high valuation buyers. At the same time, sellers on submarket $(1, 2)$ would have the same order filling speeds. Thus, equilibrium intermediation would result in a lower order filling rates and ultimately reduce welfare.

To connect this intuition with our main intermediation result, when the matching function satisfies the conditions from Proposition 4, the executions speed for both buyers and sellers are inversely related through the submarket's queue length. Thus, there is a trade-off: either the queue length is low and buyers trade fast or it is high and sellers trade fast, but not both.[21] As a result, intermediation improves welfare: the medium type investor can act as a broker and provide fast trading speeds to both the low and high type agents. When the matching function is linear, on the other hand, this trade-off disappears. For one side of the market trading speed is exogenous, hence, the medium type agents do not have the capacity to deliver a faster trading speed to these agents. Thus, intermediation cannot improve welfare. The proposition below states that this intuition generalizes to the case of I types.

Proposition 5. *If the matching function is linear in either the mass of sellers or the mass of buyers, intermediation cannot occur in equilibrium.*

In the case of the Leontief matching function, whether or not intermediation takes place in equilibrium depends on the level of asset supply. When the asset is scarce, it is efficient to move the asset to type 3 investors as quickly as possible. For asset levels below 1/3, it is so scarce that there are very few owners of types 1 and 2 and many non-owners of type 3. In that situation, it is optimal to not have types 1 and 2 trade with each other as this will pull away sellers from the market $(1, 3)$ and reduce the speed with which owners type 3 receive the asset. When the asset supply is between 1/3 and 1/2 there are relatively more owners of type 1 and 2. This makes it possible to cover the demand of type 3 investors so that $q_{2,3} < q_{1,3} = 1$ when market $(1, 2)$ does not open. In this case, type 3 non-owners receive the asset with the fastest possible speed of μ. Since type 2 sellers are relatively abundant, however, they can fully cover a little extra demand from type 3 investors if they were to switch from market $(1, 3)$ to $(2, 3)$. This makes it possible to open market $(1, 2)$ and move type 1 sellers without reducing the average speed with which type 3 buyers receive the asset. Thus, unlike the linear matching technology case, intermediation does not reduce order execution/filling

[21] This trade-off is reinforced by the equilibrium effect on the endogenous distribution of investor types. A high queue length leads to fast order execution speeds for sellers which, in turn, implies that sellers exit the market fast. This leads to a low mass of sellers in equilibrium and ultimately increases the queue length further.

rates. Similarly, when the asset is relatively abundant, the frictionless allocation features investors of both type 2 and 3 owning the asset. Hence, the efficient frictional allocation will have the asset move away from investors type 1 as quickly as possible. When the asset supply is above 2/3, there are few non-owners of type 2 and 3 and they both buy from type 1 in an attempt to cover as much of the supply as possible. When the asset is between 1/2 and 2/3, there are relatively more non-owners of types 2 and 3 and they can meet all the supply of type 1 owners, even if some non-owners of type 3 trade with type 2 owners.

To sum up, the model predictions under a Leontief matching function are similar to the linear matching case for either low or high asset supply levels. For intermediate levels of asset supply, the predictions differ and this may open the door to intermediation. Under a Leontief matching function both sides of the market contribute to congestion, but this contribution is limited as market participants can achieve the maximum order execution/filling rates when the queue length is one. Hence, intermediation may arise in equilibrium but it does not strictly imrpove welfare. This result hinges on the kink inherent in the Leontief matching function. It shows, however, that even in the case of a smoother matching function, the welfare improvement due to equilibrium intermediation may be limited.

5 Numerical Examples

The analysis so far has focused on the analytical exploration of the model's properties to illuminate the emergence of intermediation as an equilibrium outcome. In this Section, we gain further insight into the mechanics of the model by investigating some of its numerical properties. To this end, we compare our model's predictions with those from a standard OTC model with random search, as developed in Hugonnier et al. (2021). We analyze the effects of changes in matching efficiency, the dispersion of investor valuations, and the mass of intermediate valuation agents on the cost of immediacy and the volume of intermediated

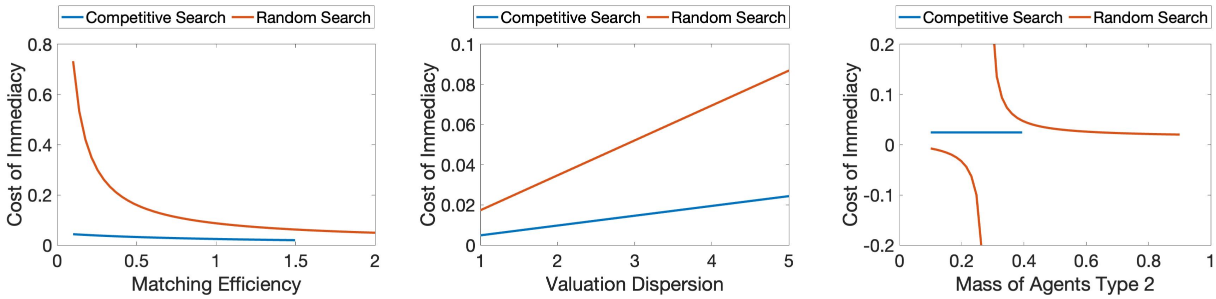


Figure 1: Cost of Immediacy in Random and Competitive Search.
Cost of immediacy for different levels of matching efficiency μ (left), valuation dispersion $|\delta_1 - \delta_2| = |\delta_3 - \delta_2|$ (middle), and the mass of type 2 agents $f(\delta_2)$ (right) in the random (red line) and competitive (blue line) search models.

trade. Our focus is on these two variables because they are studied extensively in the OTC literature and capture both price and quantity dimensions of trade. Appendix D contains a more detailed numerical exploration of the two models.

We focus on a three-type version of the two models where investor valuations are $\delta_1 = 0$, $\delta_2 = 1/2$, and $\delta_3 = 1$. Types 1 and 3 are natural customers, whereas type 2 may act as an intermediary. The discount rate is set at 5% and the arrival rate of preference shocks is $\gamma = 1$, following Hugonnier et al. (2014). The asset supply is $A = 0.5$, and the matching efficiency is $\mu = 1$.[22] The underlying distribution of utility types is assumed to be uniform and we assume equal bargaining weights for buyers and sellers in the random search model. In the competitive search model meeting rates are endogenous and formed through a matching function. We use the Cobb-Douglas specification, $m(o, n) = \mu o^{1/2} n^{1/2}$, the most commonly-used matching function in the search and matching literature, which also satisfies the Inada-type conditions of Proposition 4.

Figure 1 plots how the cost of immediacy changes for various comparative statics. Our measure of the cost of immediacy captures how much more expensive is for a customer of type 3 to buy the asset quickly from the middle-value agent type 2 (who plays the role of

[22] The matching efficiency amounts to the meeting rate in the random search case but it multiplies the endogenous meeting rate from the matching function in our model.

the intermediary) versus buying it at a lower speed from the low-value investor 1. More concretely, the cost if immediacy is defined as the percentage change of price $p(\delta_2, \delta_3)$ versus $p(\delta_1, \delta_3)$ over the percentage change in the expected order filling time from investor type 2 versus 1. Thus, our measure for the cost of immediacy summarizes the behavior of both prices and trading speeds.[23] We expect the behavior of this measure to be different in the two models, because in the competitive search framework the agents' value functions correspond to the shadow values of the planner's problem. This, however, is not true in the random search model. As a result, the cost of immediacy in the random search framework does not reflect the planner's willingness to pay for immediacy.[24]

The first comparative static (left panel of Figure 1) shows that the cost of immediacy decreases sharply as matching efficiency increases in the random search model, while it is relatively flat in our model. The reason is that under random search trading speeds are fixed as prices change, while in our model trading speeds adjust and make the impact of price changes on the cost of immediacy smaller. Our second comparative static (middle panel of Figure 1) explores the impact of changes in the dispersion of investor valuations.[25] Intuitively, in both models the cost of immediacy increases, since the type 3 investor is willing to pay more for the asset. In the random search model, this is solely due to changes in prices, since transaction rates are fixed. In the competitive search model, however, there is a trade-off between posted prices and trading speeds, which attenuates the increase in the cost of immediacy.[26] The third comparative static highlights even more the importance of

[23] We explore two alternative measures for the cost of immediacy in Appendix D.

[24] We thank an anonymous referee for drawing our attention to this connection.

[25] Specifically, we set $\delta_2 = 5$ and vary δ_1 and δ_3 in the intervals $[0, 4]$ and $[6, 10]$, such that δ_2 is always equidistant to δ_1 and δ_3.

[26] To be precise, this comparative static is a mean-preserving spread. As a result, the amount by which agents type 3 value the asset more than agents type 2 is always the same as the amount by which agents type 2 value it more than agents type 1. Consequently, the queue lengths do not change as we increase valuation dispersion. Nonetheless, when sellers post terms of trade there is still a trade-off between prices and order execution speeds which attenuates the increase in the cost of immediacy relative to the random search model.

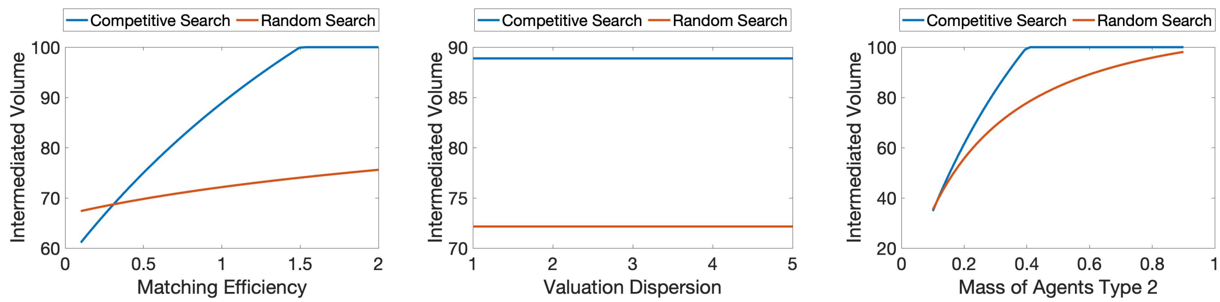


Figure 2: Intermediated Trade in Random and Competitive Search.
Percentage of trade intermediated by agents type 2 for different levels of matching efficiency μ (left), valuation dispersion $|\delta_1 - \delta_2| = |\delta_3 - \delta_2|$ (middle), and the mass of type 2 agents $f(\delta_2)$ (right) in the random (red line) and competitive (blue line) search models.

the link between prices and trading speeds for the numerical predictions of the two models. In the competitive search framework, buyers will pay higher prices only if they come with faster trading speeds, which leads to the cost of immediacy always being positive. In the random search model, however, relative trading speeds depend solely on the mass of agents, and this may make the cost of immediacy either positive or negative. For relatively large masses of agents type 2, buyers will typically receive the asset faster from them than from the type 1 customer. Since $p(\delta_1, \delta_3)$ is always smaller than $p(\delta_2, \delta_3)$, this results in a positive cost of immediacy. As the mass of type 2 agents decreases, however, their order execution edge diminishes and the cost of immediacy explodes, since going through a dealer provides only a small increase in execution speeds for a discrete increase in price. When the mass of dealers is small enough, they cannot provide speeds higher than type 1 agents, so the cost of immediacy becomes negative, i.e. type 1 customers provide both higher execution speeds and sell the asset cheaper.

The two models also differ in their implications for intermediated trade volume (the volume of transactions made by dealers as a percentage of total volume), as seen in Figure 2. When we increase either matching efficiency or the mass of dealer agents (left and right panels), intermediated volume increases in both models, but this increase is faster for the competitive search model. Intuitively, when customers can trade faster with dealers, they choose to do so and very few trades are direct customer transactions. In the extreme, when agents can trade very fast with dealers (either μ or $f(\delta_2)$ is very high), all customers choose to go through a dealer, market $(1, 3)$ closes, and dealers intermediate all trade. In the random search model, agents do not have that choice, thus customer to customer trades are more common. Although the competitive search model generally features higher intermediated trade volume than the random search one, in both models the volume does not change when we increase the dispersion of investor valuations (middle panel). This is because, in both models, a mean-preserving spread in investor valuations leaves trading speeds unchanged.[27]

6 Conclusion

In this paper, we have built a search-theoretic model of a fully decentralized OTC market with price competition. Asset owners, who are heterogeneous with respect to their asset valuation, post prices that are available to potential buyers; buyers, who are also heterogeneous in their asset valuations, observe all prices and decide to which seller they will direct their order. Our main result is that equilibrium intermediation emerges endogenously: some agents choose to act as dealers, selling the asset when they have it and buying the asset when they do not have it. Moreover, intermediation is extensive: for a broad class of matching functions satisfying an asymptotic condition, all agents, except the lowest and highest valuation ones, choose to intermediate trades.

If the matching function does not satisfy the asymptotic condition, intermediation is still likely to occur but its extent may be limited. To understand this, we explore the prevalence of intermediation in our model in a series of analytical and numerical examples.

[27] In the random search model, agents' valuations do not affect execution speeds in general. In the competitive search model, this is true in this particular case because an increase in valuation dispersion does not alter the relative amount by which investor type 3 values the asset more than investor type 2, as compared to the amount by which investor type 2 values the asset more than type 1.

Analytically, we show that intermediation is an emergent property of the market when both buyers and sellers contribute to congestion. Our numerical exercises also contrast the model's equilibrium implications to those of a benchmark model with random meetings and ex post bargaining. We find that the trade-off between prices and trading speeds, which is a feature of the competitive search but not of the random search framework, has important implications for key equilibrium objects, such as the cost of immediacy and the intermediated trade volume.

References

Acemoglu, D. and Shimer, R. (1999). Efficient unemployment insurance. *Journal of political Economy*, 107(5):893–928.

Adrian, T., Fleming, M., Shachar, O., and Vogt, E. (2017). Market liquidity after the financial crisis. *Annual Review of Financial Economics*, 9:43–83.

Afonso, G. and Lagos, R. (2015). Trade dynamics in the market for federal funds. *Econometrica*, 83(1):263–313.

Afonso, G. M. and Lagos, R. (2014). An empirical study of trade dynamics in the fed funds market. Technical report, Federal Reserve Bank of Minneapolis.

Albrecht, J., Gautier, P. A., and Vroman, S. (2016). Directed search in the housing market. *Review of Economic Dynamics*, 19:218–231.

Armenter, R. and Lester, B. (2017). Excess reserves and monetary policy implementation. *Review of Economic Dynamics*, 23:212–235.

Bessembinder, H. and Maxwell, W. (2008). Markets: Transparency and the corporate bond market. *Journal of economic perspectives*, 22(2):217–234.

Bessembinder, H., Spatt, C., and Venkataraman, K. (2020). A survey of the microstructure of fixed-income markets. *Journal of Financial and Quantitative Analysis*, 55(1):1–45.

Bethune, Z., Sultanum, B., and Trachter, N. (2019). An information-based theory of financial intermediation.

Chade, H., Eeckhout, J., and Smith, L. (2017). Sorting through search and matching models in economics. *Journal of Economic Literature*, 55(2):493–544.

Chang, B. (2017). Adverse selection and liquidity distortion. *The Review of Economic Studies*, 85(1):275–306.

Chang, B. and Zhang, S. (2018). Endogenous market making and network formation. *Available at SSRN 2600242*.

Duffie, D., Gârleanu, N., and Pedersen, L. H. (2005). Over-the-counter markets. *Econometrica*, 73(6):1815–1847.

Duffie, D., Gârleanu, N., and Pedersen, L. H. (2007). Valuation in over-the-counter markets. *The Review of Financial Studies*, 20(6):1865–1900.

Eeckhout, J. and Kircher, P. (2010). Sorting and decentralized price competition. *Econometrica*, 78(2):539–574.

Farboodi, M., Jarosch, G., and Menzio, G. (2017a). Intermediation as rent extraction. Technical report, National Bureau of Economic Research.

Farboodi, M., Jarosch, G., and Shimer, R. (2017b). The emergence of market structure. Technical report, National Bureau of Economic Research.

Gabrovski, M. and Ortego-Marti, V. (2019). The cyclical behavior of the beveridge curve in the housing market. *Journal of Economic Theory*, 181:361–381.

Gabrovski, M. and Ortego-Marti, V. (2021a). Efficiency in the housing market with search frictions. Technical report, Mimeo, University of California Riverside.

Gabrovski, M. and Ortego-Marti, V. (2021b). Search and credit frictions in the housing market. *European Economic Review*, 134:103699.

Garriga, C. and Hedlund, A. (2020). Mortgage debt, consumption, and illiquid housing markets in the great recession. *American Economic Review*, 110(6):1603–34.

Geromichalos, A. and Herrenbrueck, L. (2016). Monetary policy, asset prices, and liquidity in over-the-counter markets. *Journal of Money, Credit and Banking*, 48(1):35–79.

Guerrieri, V. and Shimer, R. (2014). Dynamic adverse selection: A theory of illiquidity, fire sales, and flight to quality. *American Economic Review*, 104(7):1875–1908.

Guerrieri, V. and Shimer, R. (2018). Markets with multidimensional private information. *American Economic Journal: Microeconomics*, 10(2):250–74.

Guerrieri, V., Shimer, R., and Wright, R. (2010). Adverse selection in competitive search equilibrium. *Econometrica*, 78(6):1823–1862.

Head, A., Lloyd-Ellis, H., and Sun, H. (2014). Search, liquidity, and the dynamics of house prices and construction. *American Economic Review*, 104(4):1172–1210.

Hollifield, B., Neklyudov, A., and Spatt, C. (2017). Bid-ask spreads, trading networks, and the pricing of securitizations. *The Review of Financial Studies*, 30(9):3048–3085.

Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *The Review of Economic Studies*, 57(2):279–298.

Hugonnier, J., Lester, B., and Weill, P.-O. (2014). Heterogeneity in decentralized asset markets. Technical report, National Bureau of Economic Research.

Hugonnier, J., Lester, B., and Weill, P.-O. (2018). Frictional intermediation in over-the-counter markets. Technical report, National Bureau of Economic Research.

Hugonnier, J., Lester, B., and Weill, P.-O. (2021). Heterogeneity in decentralized asset markets. Technical report.

Jerez, B. (2014). Competitive equilibrium with search frictions: A general equilibrium approach. *Journal of Economic Theory*, 153:252–286.

Kargar, M., Passadore, J., and Silva, D. (2020). A competitive search theory of asset pricing. *Available at SSRN*.

Lagos, R. and Rocheteau, G. (2007). Search in asset markets: Market structure, liquidity, and welfare. *American Economic Review*, 97(2):198–202.

Lagos, R. and Rocheteau, G. (2009). Liquidity in asset markets with search frictions. *Econometrica*, 77(2):403–426.

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

Lagos, R., Rocheteau, G., and Wright, R. (2017). Liquidity: A new monetarist perspective. *Journal of Economic Literature*, 55(2):371–440.

Lester, B., Rocheteau, G., and Weill, P.-O. (2015). Competing for order flow in otc markets. *Journal of Money, Credit and Banking*, 47(S2):77–126.

Li, D. and Schürhoff, N. (2019). Dealer networks. *The Journal of Finance*, 74(1):91–144.

Li, Q. (2019). Securitization and liquidity creation in markets with adverse selection. *Available at SSRN 2932441*.

Liu, Y., Vogel, S., and Zhang, Y. (2018). Electronic trading in otc markets vs. centralized exchange. In *Paris December 2018 Finance Meeting EUROFIDAI-AFFI*.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of political Economy*, 105(2):385–411.

Mortensen, D. T. and Wright, R. (2002). Competitive pricing and efficiency in search equilibrium. *International economic review*, 43(1):1–20.

Neklyudov, A. (2019). Bid-ask spreads and the over-the-counter interdealer markets: Core and peripheral dealers. *Review of Economic Dynamics*, 33:57–84.

Seierstad, A. and Sydsaeter, K. (1986). *Optimal control theory with economic applications*. Elsevier North-Holland, Inc.

Shi, S. (2001). Frictional assignment. i. efficiency. *Journal of Economic Theory*, 98(2):232–260.

Shimer, R. (2005). The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy*, 113(5):996–1025.

Stafford, P. (2016). Icap's future reflects derivatives market in transition. *Financial Times*.

Üslü, S. (2019). Pricing and liquidity in decentralized asset markets. *Econometrica*, 87(6):2079–2140.

Vogel, S. (2019). When to introduce electronic trading platforms in over-the-counter markets? *Available at SSRN 2895222*.

Weill, P.-O. (2020). The search theory of over-the-counter markets. *Annual Review of Economics*, 12:747–773.

Wheaton, W. C. (1990). Vacancy, search, and prices in a housing market matching model. *Journal of political Economy*, 98(6):1270–1292.

Williams, B. (2019). Search, liquidity, and retention: Screening multidimensional private information.

Wright, R., Kircher, P., Julien, B., and Guerrieri, V. (2019). Directed search and competitive search: A guided tour. *Journal of Economic Literature*.

A Equilibrium Existence and Constrained Efficiency

Here we establish existence and constrained efficiency of the decentralized equilibrium. To do so, we show that the equilibrium of the decentralized model coincides with the solution to the planner's problem. The complete planner's problem is presented below. All variables are functions of time, but we do not make that explicit in the notation to keep it succinct.

$$\max_{\{o(\delta_i,\delta_j)\}_{i,j\in\mathcal{I}},\{n(\delta_i,\delta_j)\}_{i,j\in\mathcal{I}\in U}} \int_0^\infty exp(-rt)\sum_i o(\delta_i)\delta_i dt \tag{14}$$

s.t.

$$\begin{aligned} \dot{o(\delta_i)} = \gamma f(\delta_i)\sum_k o(\delta_k) - \gamma o(\delta_i)\sum_k f(\delta_k) + \sum_k m\left(o(\delta_k,\delta_i), n(\delta_k,\delta_i)\right) \\ - \sum_j m\left(o(\delta_i,\delta_j), n(\delta_i,\delta_j)\right), \quad \forall i \end{aligned} \tag{15}$$

$$\begin{aligned} \dot{n(\delta_j)} = \gamma f(\delta_j)\sum_k n(\delta_k) - \gamma n(\delta_j)\sum_k f(\delta_k) + \sum_k m\left(o(\delta_j,\delta_k), n(\delta_j,\delta_k)\right) \\ - \sum_i m\left(o(\delta_i,\delta_j), n(\delta_i,\delta_j)\right), \quad \forall j \end{aligned} \tag{16}$$

$$o(\delta_i) - \sum_j o(\delta_i,\delta_j) \geq 0, \quad \forall i \tag{17}$$

$$n(\delta_j) - \sum_i n(\delta_i,\delta_j) \geq 0, \quad \forall j \tag{18}$$

$$o(\delta_i,\delta_j) \geq 0, \quad \forall i,j \tag{19}$$

$$n(\delta_i,\delta_j) \geq 0, \quad \forall i,j \tag{20}$$

$$\lim_{t\to\infty} o(\delta_i) \geq 0, \quad \lim_{t\to\infty} n(\delta_i) \geq 0, \quad \forall i \tag{21}$$

$$o(\delta_i),\ n(\delta_i) \text{ given at time } t=0, \quad \forall i \tag{22}$$

where $U = [0,1]^{I \times I \times 2}$ is the range over which the planner optimizes the masses of owners and non-owners.

The planner's objective is to maximize the net present sum of the instantaneous payoffs of all owners. She accomplishes this by assigning owners and non-owners as buyers and sellers in different submarkets. We assume that if the planner finds it optimal for owner type i to trade with non-owner type j, then all such trades take place in a single market. This is without loss of generality, since the matching function has constant returns to scale and meetings within a submarket are random.[28] The planner essentially moves masses of owners and non-owners around different submarkets to maximize the total surplus of potential matches. When considering these moves, the planner is constrained by the matching function in each submarket, as well as the laws of motion that determine the evolution of investor masses over time. Of course, the planner is also constrained by the appropriate non-negativity and resource constraints.

Analyzing the planner's problem allows us to derive two important results: first, the solution to the planner's problem coincides with the decentralized equilibrium. Second, a solution to the planner's problem (and, as a result, a decentralized equilibrium) exists. This is our Proposition 6 below. Although OTC search models with exogenous contact rates are typically contrained efficient (see, e.g., Hugonnier et al. (2021) or Afonso and Lagos (2015)), this does not trivially generalize in the case of endogenous contact rates. Specifically, Lagos and Rocheteau (2007) and Lester et al. (2015) show that in a model with random search, ex post bargaining and contact rates given by a matching function, the equilibrium is typically inefficient, unless a version of the Hosios (1990) condition holds. The equilibrium is also inefficient in the model of Farboodi et al. (2017b), in which contact rates are the result of ex-ante identical agents choosing how much to invest in the quality of their search technology.

[28] Further, we assume that if the planner is indifferent between assigning (non-)owners type i, j to the same submarket and to different submarkets, she will assign them to different ones. This is also without loss of generality due to the constant returns to scale assumption.

These inefficiencies arise due to externalities in the matching process and ex post bargaining between investors. As is standard in the competitive search literature, we show that when owners post prices and non-owners choose which price to search for, these externalities are internalized through this pricing mechanism. In this sense, our Proposition 6 generalizes the results of Lester et al. (2015) for a fully decentralized market.

Proposition 6. *The decentralized equilibrium exists and it is constrained efficient.*

Proof. **Existence.** First we show that a solution of the planner's problem exists. To this end, we will invoke Theorem 21 in Seierstad and Sydsaeter (1986) on p.406. To show that our problem satisfies the conditions in the theorem, we translate our problem into their notation and then show each of the 5 conditions hold. First, let $x := (\{o(\delta_i)\}_{i\in\mathcal{I}}, \{n(\delta_j)\}_{j\in\mathcal{I}})$ be the vector of state variables, $u := (\{o(\delta_i, \delta_j)\}_{i,j\in\mathcal{I}}, \{n(\delta_i, \delta_j)\}_{i,j\in\mathcal{I}})$ be the vector of market allocations, $f_0(x, u, t) := exp(-rt)\sum_i o(\delta_i)\delta_i$ be the maximand, $f(x, u)$ be a vector function such that for $i \in [1, I]$, $f_i(x, u)$ is the right hand side of equation (15) for owner type i and for $j \in [I + 1, 2I]$, $f_j(x, u)$ is the right hand side of equation (16) for non-owner type j. Similarly, define $g(x, u)$ to be the left hand side of the constraints (17), (18).

Claim 1. The functions f_0, f, and g are continuos and U is closed. This is obvious.

Claim 2. Define $N(x, U, t) = \{(f_0(x, u, t) + \tilde{\gamma}, f(x, u)) : \tilde{\gamma} \leq 0, g_i(x, u, t) \geq 0, \forall i, u \in U\}$. We will show that this set is convex for all $(x, t) \in \mathbb{R}^{2I} \times [0, \infty)$. Pick any (x, t) and observe that the set $N(x, U, t)$ is all pairs $(f_0(x, u, t) + \tilde{\gamma}, f(x, u))$ induced by some $\tilde{\gamma} \leq 0$ and some $u \in U$ for which the constraints (17), (18) are satisfied. Then, let $(f_0(x, u', t) + \tilde{\gamma}', f(x, u'))$ and $(f_0(x, u'', t) + \tilde{\gamma}', f(x, u''))$ be some elements of $N(x, U, t)$ and let $\lambda \in (0, 1)$. Then, notice that $\lambda(f_0(x, u', t) + \tilde{\gamma}') + (1 - \lambda)(f_0(x, u'', t) + \tilde{\gamma}') = f_0(x, u', t) + \lambda\tilde{\gamma}' + (1 - \lambda)\tilde{\gamma}''$, with $\lambda\tilde{\gamma}' + (1 - \lambda)\tilde{\gamma}'' \leq 0$. Next, take any $i, j \in \mathcal{I}$ and observe that by the concavity of the matching function,

$$\lambda m\left(o'(\delta_k,\delta_i),n'(\delta_k,\delta_i)\right)+(1-\lambda)m\left(o''(\delta_k,\delta_i),n''(\delta_k,\delta_i)\right)\leq$$
$$m\left(\lambda o'(\delta_k,\delta_i)+(1-\lambda)o''(\delta_k,\delta_i),\lambda n'(\delta_k,\delta_i)+(1-\lambda)n''(\delta_k,\delta_i)\right). \tag{23}$$

Then, by continuity and strict monotonicity of $m(\cdot)$ there exists some tuple $(\tilde{o}(\delta_i,\delta_j),\tilde{n}(\delta_i,\delta_j))\leq(\lambda o'(\delta_i,\delta_j)+(1-\lambda)o''(\delta_i,\delta_j),\lambda n'(\delta_i,\delta_j)+(1-\lambda)n''(\delta_i,\delta_j))$, such that

$$m(\tilde{o}(\delta_k,\delta_i),\tilde{n}(\delta_k,\delta_i))=\lambda m\left(o'(\delta_k,\delta_i),n'(\delta_k,\delta_i)\right)+(1-\lambda)m\left(o''(\delta_k,\delta_i),n''(\delta_k,\delta_i)\right). \tag{24}$$

Then, let $\tilde{u}$ be a tuple of market allocations such that (24) holds for all $i,j\in\mathcal{I}$. It is easy to see that $\tilde{u}\in U$ and $f(x,\tilde{u},t)=\lambda f(x,u',t)+(1-\lambda)f(x,u'',t)$. Lastly, we need to show that $\tilde{u}$ is such that $g(x,\tilde{u},t)\geq 0$. But this clearly holds since

$$\sum_j\tilde{o}(\delta_i,\delta_j)\leq\sum_j\left[\lambda o'(\delta_i,\delta_i)+(1-\lambda)o''(\delta_i,\delta_i)\right]\leq o(\delta_i),\quad\forall i \tag{25}$$
$$\sum_i\tilde{n}(\delta_i,\delta_j)\leq\sum_i\left[\lambda n'(\delta_i,\delta_i)+(1-\lambda)n''(\delta_i,\delta_i)\right]\leq n(\delta_j),\quad\forall j \tag{26}$$

Thus, $\tilde{u}\in N(x,U,t)$, which proves the claim.

Claim 3. Let $\Gamma=\{(x,u,t):g(x,u,t)\geq 0,u\in U,t\in[0,\infty)\}$ and $\Gamma_t=\{x:(x,u,t)\in\Gamma$ for some $u\in U\}$. If $x_n\in\Gamma_t$, $v_n\in N(x_n,U,t)$, $x_n\to x$, $v_n\to v$, then $x\in\Gamma_t$ and $v\in N(x,U,t)$. Observe that any tuple of state variables that satisfies the aggregate resource constraints $\sum_i o(\delta_i)=A$, $\sum_j n(\delta_j)=1-A$ is an element of Γ_t. Since all x_n satisfy those, so does x and hence $x\in\Gamma_t$. Next, let $f(x_n,u_n,t)$ be the vector induced by x_n, v_n. By continuity of the matching function, $f(x_n,u_n,t)\to f(x,u,t)$, where $f(x,u,t)$ is the vector induced by x, v. Next, it is easy to see that $f_0(x_n,t)\to f_0(x,t)$. Thus, all we are left to do to prove the claim is to show that $u\in U$ and that $g(x,u,t)\geq 0$. Since U is closed the former holds. Since $g(x_n, u_n, t) \geq 0$ for all n, then the latter holds as well. Thus, $v \in N(x, U, t)$.

Claim 4. For each $p \neq 0$ there exist locally integrable functions $\varphi_p(t)$ and $\psi_p(t)$ such that for all $(x, u, t) \in \Gamma$, $f_0(x, u, t) + p \cdot f(x, u, t) \leq \varphi_p(t) + \psi_p(t)||x||$ and there exists a constant M such that $||x(t_0)|| \leq M$ for all admissible x and $f_0(x, u, t) \leq \hat{\gamma}(t)$ for all $(x, u, t) \in \Gamma$. Then, it is easy to see that $\sum_i \delta_i > exp(-rt) \sum_i o(\delta_i)\delta_i$, since $o(\delta_i) \leq A < 1$. Next, observe that for $i \in [1, I]$ and $j \in [I+1, 2I]$,

$$f_i(x, u, t) \leq \gamma f(\delta_i)A + \sum_k m(o(\delta_k, \delta_i), n(\delta_k, \delta_i) \leq \gamma A + Im(1,1) \leq \gamma + Im(1,1), \tag{27}$$

$$f_j(x, u, t) \leq \gamma f(\delta_{j-I})(1-A) + \sum_k m(o(\delta_{j-I}, \delta_k), n(\delta_{j-I}, \delta_k) \leq \gamma(1-A) + Im(1,1) \leq \gamma + Im(1,1). \tag{28}$$

Then, let $\varphi_p(t) = \sum_i \delta_i + p \cdot \mathbf{M(1,1)}$, where $\mathbf{M(1,1)}$ is a vector of length $2I$ whose elements are $\gamma + Im(1,1)$. Then, set $\psi_p(t) = 0$ and observe that by construction $f_0(x, u, t) + p \cdot f(x, u, t) \leq \varphi_p(t) + \psi_p(t)||x||$, with both $\varphi_p(t)$ and $\psi_p(t)$ both locally integrable. Next, observe that there is a unit mass of agents in the economy and so $1 \geq ||x||$. If $\hat{\gamma}(t) = \sum_i \delta_i$, clearly $\hat{\gamma}(t) \geq f_0(x, u, t)$, for all $(x, u, t) \in \Gamma$. This proves the claim.

Claim 5. There exist integrable functions $v^i(t)$ defined on $[0, \infty)$ such that for all admissible tuples (x, u), for all t, $f_i(x, u) \leq v^i(t)$. Take $v^i(t) = \gamma + Im(1,1)$. From the preceding claim, it follows that $f_i(x, u) \leq v^i(t)$ for all t. Lastly, $v^i(t)$ is integrable, so the claim holds.

Together claims 1 through 5 establish that our problem satisfies the conditions outlined in Theorem 21 on page 406 from Seierstad and Sydsaeter (1986). Thus, an optimal tuple (x, u) exists.

Characterizing Planner's Allocation. Define the Hamiltonian for the original planner's problem by

$$
\begin{aligned}
H \equiv & \sum_i o(\delta_i)\delta_i \\
& + \sum_i \tilde{\lambda}_i \left[\gamma f(\delta_i) \sum_k o(\delta_k) - \gamma o(\delta_i) + \sum_k m\left(o(\delta_k, \delta_i), n(\delta_k, \delta_i)\right) - \sum_j m\left(o(\delta_i, \delta_j), n(\delta_i, \delta_j)\right) \right] \\
& + \sum_j h_j \left[\gamma f(\delta_j) \sum_k n(\delta_k) - \gamma n(\delta_j) + \sum_k m\left(o(\delta_j, \delta_k), n(\delta_j, \delta_k)\right) - \sum_i m\left(o(\delta_i, \delta_j), n(\delta_i, \delta_j)\right) \right],
\end{aligned}
\tag{29}
$$

where $\tilde{\lambda}_i$ is the co-state associated with law of motion for $o(\delta_i)$ and h_j is the co-state associated with the low of motion for $n(\delta_j)$. Thus, the Lagrangian of the problem is given by

$$
\begin{aligned}
L \equiv & H + \sum_i \mu_i \left(o(\delta_i) - \sum_j o(\delta_i, \delta_j) \right) + \sum_j \nu_j \left(n(\delta_j) - \sum_i n(\delta_i, \delta_j) \right) \\
& + \sum_i \sum_j \tilde{o}_{ij} o(\delta_i, \delta_j) + \sum_i \sum_j \tilde{n}_{ij} n(\delta_i, \delta_j),
\end{aligned}
\tag{30}
$$

where μ_i, ν_j, $\tilde{o}_{ij}$, $\tilde{n}_{ij}$ are the associated Lagrange multipliers. Then, the first order conditions for optimality are given by

$$\frac{\partial L}{\partial o(\delta_i)} = r\tilde{\lambda}_i - \dot{\tilde{\lambda}}_i, \quad \forall i \tag{31}$$

$$\frac{\partial L}{\partial n(\delta_j)} = rh_j - \dot{h}_j, \quad \forall j \tag{32}$$

$$\frac{\partial L}{\partial o(\delta_i, \delta_j)} = 0, \quad \forall i, j \tag{33}$$

$$\frac{\partial L}{\partial n(\delta_i, \delta_j)} = 0, \quad \forall i, j \tag{34}$$

$$\mu_i \left(o(\delta_i) - \sum_j o(\delta_i, \delta_j) \right) = 0, \quad \forall i \tag{35}$$

$$\nu_j \left(n(\delta_j) - \sum_i n(\delta_i, \delta_j) \right) = 0, \quad \forall j \tag{36}$$

$$\tilde{o}_{ij} o(\delta_i, \delta_j) = 0, \quad \forall i, j \tag{37}$$

$$\tilde{n}_{ij} n(\delta_i, \delta_j) = 0, \quad \forall i, j \tag{38}$$

$$\lim_{t \to \infty} exp(-rt)\tilde{\lambda}_i o(\delta_i) = 0, \quad \forall i \tag{39}$$

$$\lim_{t \to \infty} exp(-rt) h_j n(\delta_j) = 0, \quad \forall j \tag{40}$$

We can reduce the first four first order conditions to the following system

$$(r + \gamma)\tilde{\lambda}_i = \delta_i + \gamma \mathbb{E}\tilde{\lambda} + \mu_i + \dot{\tilde{\lambda}}, \tag{41}$$

$$(r + \gamma) h_j = \gamma \mathbb{E} h + \nu_j + \dot{h}, \tag{42}$$

$$\mu_i = \lambda(q(\delta_i, \delta_j))\eta(q(\delta_i, \delta_j)) \left[\tilde{\lambda}_j - h_j - (\tilde{\lambda}_i - h_i) \right] + \tilde{o}(\delta_i, \delta_j), \tag{43}$$

$$\nu_j = \frac{\lambda(q(\delta_i, \delta_j))}{q(\delta_i, \delta_j)} \left[1 - \eta(q(\delta_i, \delta_j)) \right] \left[\tilde{\lambda}_j - h_j - (\tilde{\lambda}_i - h_i) \right] + \tilde{n}(\delta_i, \delta_j), \tag{44}$$

where $\mathbb{E}\tilde{\lambda}$ and $\mathbb{E}h$ are the average $\tilde{\lambda}_i$ and h_j. Then, noting that in steady state the values of

$\tilde{\lambda}_i$ and h_j are constant, it follows that we can reduce the system above to

$$(r+\gamma)\tilde{\lambda}_i = \delta_i + \gamma\mathbb{E}\tilde{\lambda} + \lambda(q(\delta_i,\delta_j))\eta(q(\delta_i,\delta_j))\left[\tilde{\lambda}_j - h_j - (\tilde{\lambda}_i - h_i)\right] + \tilde{o}(\delta_i,\delta_j), \tag{45}$$

$$(r+\gamma)h_j = \gamma\mathbb{E}h + \frac{\lambda(q(\delta_i,\delta_j))}{q(\delta_i,\delta_j)}\left[1-\eta(q(\delta_i,\delta_j))\right]\left[\tilde{\lambda}_j - h_j - (\tilde{\lambda}_i - h_i)\right] + \tilde{n}(\delta_i,\delta_j). \tag{46}$$

With appropriate relabeling, these conditions are exactly the same as the Bellman equations along the decentralized equilibrium, (8) and (9), when $\tilde{o}(\delta_i,\delta_j) = \tilde{n}(\delta_i,\delta_j) = 0$. But, the multipliers are zero if and only if the planner assigns positive measures of owners and non-owners on this market. That is, the planner's conditions coincide with the decentralized ones along the equilibrium path of trade only when the planner assigns positive measures on that market. If there are no owners assigned on that market, this means that $\tilde{o}(\delta_i,\delta_j) > 0$ and so the value of an owner participating in that market is not large enough to yield her optimal value, i.e. the owner can do better (is more socially beneficial) if she is on a different market. Analogously for non-owners: if $\tilde{n}(\delta_i,\delta_j) > 0$ there are no non-owners on that market and the socially optimal value h_j is higher than any potential payoff that the non-owner might earn from participating that market, i.e. the non-owner cannot be compensated with her market utility from participating in the market. Thus, the decentralized equilibrium is efficient.

Sufficiency of the First Order Conditions. Next, observe that the maximand as well as constraints are continuously differentiable. Let $A(t)$ be the set of all tuples of owners, $o(\delta_i,\delta_j)$ and non-owners, $n(\delta_i,\delta_j)$, assigned to each market such that constraints (17), (18), (19), (20) are satisfied given the masses of owners, $o(\delta_i)$, and non-owners, $n(\delta_j)$, at time t. Thus, it is easy to see that this set is convex for all t. Let $\hat{x}(t)$ be a tuple of owner masses, $o(\delta_i)$, non-owner masses, $n(\delta_j)$, and owner and non-owner market participation assignments, $o(\delta_i,\delta_j)$, $n(\delta_i,\delta_j)$ such that the constraints of the problem and the first order conditions are satisfied. Given the resulting co-states $\hat{\lambda}_i$ and h_j, define $\hat{H}$ to be the maximized Hamiltonian, i.e. $\hat{H} \equiv \max_{o(\delta_i,\delta_j),o(\delta_i,\delta_j)\in A(t)} H$. Since any tuple of owner and non-owner market assignments which is an element of $\arg\max_{o(\delta_i,\delta_j),o(\delta_i,\delta_j)\in A(t)} H$ is independent of the state variables $o(\delta_i)$, $n(\delta_j)$, it is easy to see that the resulting maximized Hamiltonian is linear in the state variables and as a consequence jointly concave in them. Hence, by Arrow's sufficiency theorem $\hat{x}(t)$ is a global maximum.

■

B Proofs

Lemma 1. *For any $i > j$, the following inequalities hold: $\Delta V^*(\delta_i) > \Delta V^*(\delta_j)$; $V_0^*(\delta_i) \geq V_0^*(\delta_j)$; $V_1^*(\delta_i) > V_1^*(\delta_j)$. Furthermore, if some non-owners of type j are active in some submarket, then $V_0^*(\delta_i) > V_0^*(\delta_j)$.*

Proof. We prove the statements in the order the appear in the text. First, let $(p^*(\delta_i), q^*(\delta_i))$ be an optimal price and queue length for an owner type $o(\delta_i)$. Analogously, let $(p^{**}(\delta_i), q^{**}(\delta_i))$ be the price and queue length on a market that non-owner type $n(\delta_i)$ optimally chooses to visit. Then,

$$(r+\gamma)V_1^*(\delta_i) = \delta_i + \gamma V_1^* + \lambda(q^*(\delta_i))\left[p^*(\delta_i) - \Delta V^*(\delta_i)\right], \tag{47}$$

$$(r+\gamma)V_0^*(\delta_i) = \gamma V_0^* + \frac{\lambda(q^{**}(\delta_i))}{q^{**}(\delta_i)}\left[\Delta V^*(\delta_i) - p^{**}(\delta_i)\right]. \tag{48}$$

Now, to the contrary, suppose that there exist some types i, j, with $i > j$, such that $\Delta V^*(\delta_i) \leq \Delta V^*(\delta_j)$. Then,

$$\begin{aligned}(r+\gamma)V_0^*(\delta_j) &\geq \gamma V_0^* + \frac{\lambda(q^{**}(\delta_i))}{q^{**}(\delta_i)}\left[\Delta V^*(\delta_j) - p^{**}(\delta_i)\right] \geq \\ &\geq \gamma V_0^* + \frac{\lambda(q^{**}(\delta_i))}{q^{**}(\delta_i)}\left[\Delta V^*(\delta_i) - p^{**}(\delta_i)\right] = (r+\gamma)V_0^*(\delta_i)\end{aligned}$$

where the first inequality follows by the optimality of $(p^{**}(\delta_i), q^{**}(\delta_i))$ and the second one by assumption. Hence, $V_0^*(\delta_j) \geq V_0^*(\delta_i)$. Furthermore, $V_1^*(\delta_i) > V_1^*(\delta_j)$, since

$$\begin{aligned}(r+\gamma)V_1^*(\delta_i) &\geq \delta_i + \gamma V_1^* + \lambda(q^*(\delta_j))\left[p^*(\delta_j) - \Delta V^*(\delta_j)\right] > \\ &> \delta_j + \gamma V_1^* + \lambda(q^*(\delta_j))\left[p^*(\delta_j) - \Delta V^*(\delta_j)\right] = (r+\gamma)V_1^*(\delta_j)\end{aligned}$$

But this implies that $\Delta V^*(\delta_i) > \Delta V^*(\delta_j)$ which is a contradiction.

Next, to prove the second statement

$$(r+\gamma)V_0^*(\delta_i) \geq \gamma V_0^* + \frac{\lambda(q^{**}(\delta_j))}{q^{**}(\delta_j)}\left[\Delta V^*(\delta_i) - p^{**}(\delta_j)\right] \geq$$
$$\geq \gamma V_0^* + \frac{\lambda(q^{**}(\delta_j))}{q^{**}(\delta_j)}\left[\Delta V^*(\delta_j) - p^{**}(\delta_j)\right] = (r+\gamma)V_0^*(\delta_j)$$

The third statement follows, since $\Delta V^*(\delta_i) = V_1^*(\delta_i) - V_0^*(\delta_i) > V_1^*(\delta_j) - V_0^*(\delta_j) = \Delta V^*(\delta_j)$. Hence, $V_1^*(\delta_i) - V_1^*(\delta_j) > V_0^*(\delta_i) - V_0^*(\delta_j) \geq 0$ and $V_1^*(\delta_i) > V_1^*(\delta_j)$.

■

Proposition 1. *Every submarket in equilibrium features only one type of owner and one type of non-owner. Furthermore, if owner i and non-owner j trade in equilibrium, they trade in only one submarket.*

Proof. Consider the owner's problem given by (7). Use equations (8) and (9) along with the constraint $V_0(\delta, p) = V_0^*(\delta)$ to substitute out for the price. Then the owner's problem can be rewritten as

$$\max_{q,\delta} \{\lambda(q)S(\delta_i, \delta) - q\left[(r+\gamma)V_0^*(\delta) - \gamma V_0^*\right]\} \tag{49}$$

where $S(\delta_i, \delta) = \Delta V^*(\delta) - V_1(\delta_i, p) + V_0^*(\delta_i) = \Delta V^*(\delta) - \Delta V(\delta_i)$. Thus, each owner type chooses which non-owner types she wishes to attract and at what queue length.

We proceed by introducing three lemmas. First, we show that if a particular owner type i wishes to attract any two distinct non-owner types j and k, she will find it optimal to do so in different submarkets. Second, we show that if an owner type i wishes to attract a particular non-owner type j, there is a unique pay-off maximizing pair (p, q). Lastly, we show that no two owner types find it optimal to post the same price in equilibrium.

Lemma 4. *If an owner type i wishes to attract non-owners j and k, with $\delta_j \neq \delta_k$, she finds it optimal to do so at different prices.*

Proof. The first order condition with respect to the queue length is given by

$$\lambda'(q)S(\delta_i, \delta) = (r + \gamma)V_0^*(\delta) - \gamma V_0^* \tag{50}$$

which is equivalent to

$$\lambda'(q)S(\delta_i, \delta) = \frac{\lambda(q)}{q}\left[\Delta V^*(\delta) - p\right] \tag{51}$$

Hence, if $\varepsilon(q) \equiv \frac{\lambda'(q)q}{\lambda(q)}$ is the elasticity of the owner meeting rate $\lambda(q)$ with respect to the queue length, it follows that

$$p = [1 - \varepsilon(q)]\Delta V^*(\delta) + \varepsilon(q)\left[V_1(\delta_i, p) - V_0^*(\delta_i)\right] \tag{52}$$

Notice that there is a direct relationship between $\varepsilon(q)$ and $\eta(q)$: $\eta(q) = 1 - \varepsilon(q)$, since $\lambda(q) = m(1, q)$. But by Lemma 1 the above can hold for at most one type of non-owner δ, because ΔV^* is strictly increasing. This proves the lemma.

■

Lemma 5. *All owners type i wishing to attract non-owner type δ post the same price.*

Proof. By equation (52), the price is determined uniquely by the reservation values of the owner and non-owner which then implies that the price is determined by the queue length. Thus, conditional on attracting non-owner type δ, if owner i finds it optimal to induce a unique queue length, the lemma is proven; i.e. we need to show that equation (50) has a unique solution for q. Consider owner type i who wishes to attract non-owner of type δ. Substituting the first order condition for the price, (52), into the value functions for the owner and the non-owner, yields

$$(r+\gamma)V_1(\delta_i) = \delta_i + \gamma V_1^* + \lambda(q)\,(1-\varepsilon(q))\,S(\delta_i,\delta), \tag{53}$$

$$(r+\gamma)V_0(\delta) = \gamma V_0^* + \frac{\lambda(q)}{q}\varepsilon(q)S(\delta_i,\delta). \tag{54}$$

Hence,

$$(r+\gamma)S(\delta_i,\delta) = (r+\gamma)\Delta V^*(\delta) - \delta_i - \gamma V_1^* - \lambda(q)\,(1-\varepsilon(q))\,S(\delta_i,\delta) + (r+\gamma)V_0^*(\delta_i) \tag{55}$$

$$\Rightarrow S(\delta_i,\delta) = \frac{(r+\gamma)\,[\Delta V^*(\delta) + V_0^*(\delta_i)] - \delta - \gamma V_1^*}{r+\gamma+\lambda(q)\,(1-\varepsilon(q))} \tag{56}$$

By the concavity of $\lambda(q)$, the expression $\lambda(q)\,(1-\varepsilon(q)) = \lambda(q) - \lambda'(q)q$ is strictly increasing. Thus, both surplus and $\lambda'(q)$ are strictly decreasing in q, which then proves the lemma.

■

Thus, lemmas 4 and 5 imply that the owner's problem reduces to choosing an optimal pair (q,p) subject to attracting a non-owner type j and then choosing the non-owner type j which maximizes her payoff.

Lemma 6. *No two owners i and j, with $\delta_i \neq \delta_j$, find it optimal to post the same price given that this price attracts a positive queue length.*

Proof. Suppose not. Then there exist two owners of type i and j, with $\delta_i \neq \delta_j$, that find it optimal to post the same price p which attracts a positive queue length of non-owners, q. By Lemma 4 each owner type has a particular non-owner type they wish to attract, say a and b respectively. Since both owner types post the same price, the market is populated by both types of owners and both types of non-owners. Thus owner type i expects to trade with positive probability with both non-owner types a and b. Now consider a deviation by an owner type i who wishes to attract b on some submarket. By lemmas 4 and 5 the unique optimal price for such a deviation is $p_j \neq p$, which induces a queue length $q_j \neq q$. Since the original submarket (p, q) was a feasible option, the deviant would find it strictly better to trade with non-owner type b on a submarket p_j than on submarket p. Furthermore, market p is populated by non-owner types a and b, so the payoff of owner i on market p is the same whether she trades with type a or b. Hence, owners i posting a price p with the aim of attracting non-owners a have a strictly profitable deviation by posting a price p_j and attracting non-owner type b. Hence, price p cannot be an optimally set price for owner i.

■

This concludes the proof of Proposition 1.

■

Proposition 2. *In equilibrium, no market (δ_i, δ_j) with $i > j$ opens. Furthermore, if submarket (δ_i, δ_j) with $i < j$ opens, then markets (δ_i, δ_i) and (δ_j, δ_j) do not open.*

Proof. First, we prove the statement for $i > j$. By Lemma 1 if $i > j \Rightarrow \Delta V^*(\delta_i) > \Delta V^*(\delta_j)$. Hence, $S^*(\delta_i, \delta_j) = \Delta V^*(\delta_j) - \Delta V^*(\delta_i) < 0$. This then contradicts the optimality of (p, q) on that market since the owner would be strictly better off by posting a price that does not attract any non-owners.

Second, we prove the statement for $i < j$. Since market (δ_i, δ_j) opens, it follows that $S^*(\delta_i, \delta_j) > 0$. Suppose to the contrary that (δ_j, δ_j) opens. Then, $V_0(\delta_i, \delta_j) = V_0(\delta_j, \delta_j) = V_0^*(\delta_j)$. But since $S^*(\delta_j, \delta_j) = 0$, it follows that $\lambda'(q(\delta_i, \delta_j))\, S^*(\delta_i, \delta_j) = 0$ by equations (9) and (52). But this is a contradiction since $\Delta V^*(\delta_j) > \Delta V^*(\delta_i)$ by Lemma 1 and in equilibrium $q(\delta_i, \delta_j) = n(\delta_i, \delta_j)/o(\delta_i, \delta_j) < \infty$.

Analogously, suppose that market (δ_i, δ_i) opens. Thus, equations (8) and (52) imply that $[1 - \varepsilon(q(\delta_i, \delta_i))]\, \lambda(q(\delta_i, \delta_i))\, S^*(\delta_i, \delta_i) = [1 - \varepsilon(q(\delta_i, \delta_j))]\, \lambda(q(\delta_i, \delta_j))\, S^*(\delta_i, \delta_j)$, which is a contradiction since $S^*(\delta_i, \delta_i) = 0$ and $q(\delta_i, \delta_i) < \infty$ in equilibrium. ■

Proposition 3. *In equilibrium, an owner type i is an active seller if and only if $i \leq s$ and a non-owner type j is an active buyer if and only if $j \geq b$, where s and b are some types between 1 and I.*

Proof. First we show that an owner type i is an active seller if and only if $\delta_i \leq \delta_s$. Since there are potential gains from trade and there are no participation costs, it is easy to see that at least one market would open in equilibrium. Then, let s be the highest type among any active sellers in equilibrium. We will show that any owner $i \leq s$ is an active seller as well. Suppose to the contrary that there exists some owner type $i \leq s$ that is not an active seller. Let n denote a non-owner type that buys from the active seller type s, then

$$(r+\gamma)V_1^*(\delta_s) = \delta_s + \gamma V_1^* + \lambda(q(\delta_s,\delta_n))\left[p(\delta_s,\delta_n) - \Delta V^*(\delta_s)\right], \tag{57}$$

$$(r+\gamma)V_1^*(\delta_i) = \delta_i + \gamma V_1^*. \tag{58}$$

Since $p(\delta_s,\delta_n) - \Delta V^*(\delta_s) > 0$ and $\Delta V^*(\delta_s) \geq \Delta V^*(\delta_i)$, this means that the owner type i has a strictly profitable deviation by posting the price $p(\delta_s,\delta_n)$ and attracting length $q(\delta_s,\delta_n)$.

Similarly, let b be the lowest type among any active buyers in equilibrium. Suppose to the contrary that some non-owner type $j \geq b$ is not an active buyer. If the non-owner b trades with some owner k in equilibrium, then

$$(r+\gamma)V_0^*(\delta_b) = \gamma V_1^* + \frac{\lambda(q(\delta_k,\delta_b))}{q(\delta_k,\delta_b)}\left[\Delta V^*(\delta_b) - p(\delta_k,\delta_b)\right], \tag{59}$$

$$(r+\gamma)V_0^*(\delta_j) = \gamma V_1^*. \tag{60}$$

Since $\Delta V^*(\delta_b) - p(\delta_k,\delta_b) > 0$ and $\Delta V^*(\delta_j) \geq V^*(\delta_b)$, the non-owner type j has a strictly profitable deviation by queuing on the market (δ_k,δ_b).

■

Proposition 4. *If the matching function satisfies the Inada-type conditions* $\lim_{q\to\infty}\lambda(q) =$

$\lim_{q\to 0}\lambda(q)/q = \infty$, *then* $b = 2$ *and* $s = I - 1$.

Proof. Consider three owner types i, j, k with $\delta_i < \delta_j < \delta_k$. First, we investigate the conditions under which a given non-owner type j accepts terms of trade favourable to an owner type $i < j$ so that j is not an idle non-owner. To this end, consider some non-owner of type j and suppose that she is idle in equilibrium. Given our previous results, this means that all non-owners type j are idle. This implies that $V_0^*(\delta_j) = \gamma V_0^*/(r+\gamma)$. Then consider a deviation by some owner type $i < j$. Given that non-owners j are idle, the deviant can post a price that extracts all of the surplus and still expect a positive queue length. Then the payoff of the deviant is given by

$$V_1(\delta_i|\delta_j) = \frac{\delta_i + \gamma V_1^*}{r+\gamma} + \frac{\lambda(q)}{r+\gamma} S(\delta_i, \delta_j), \tag{61}$$

where q is the anticipated queue length and $V_1(\delta_i|\delta_j)$ is the utility of the deviant and the trade surplus is then given by $S(\delta_i,\delta_j) = \Delta V^*(\delta_j) - \Delta V(\delta_i|\delta_j)$, with $\Delta V(\delta_i|\delta_j) = V_1(\delta_i|\delta_j) - V_0^*(\delta_i)$ being the reservation value for the deviant seller. Using (61), we can express the surplus as

$$\begin{aligned} S(\delta_i,\delta_j) &= \frac{r+\gamma}{r+\gamma+\lambda(q)}\left[\Delta V^*(\delta_j) + V_0^*(\delta_i) - \frac{\delta_i + \gamma V_1^*}{r+\gamma}\right] \\ &= \frac{r+\gamma}{r+\gamma+\lambda(q)}\left[\Delta V^*(\delta_j) - \Delta V^*(\delta_i) + \frac{\lambda(q^*(\delta_i,\delta_k))\eta(q^*(\delta_i,\delta_k))}{r+\gamma} S^*(\delta_i,\delta_k)\right], \end{aligned} \tag{62}$$

where the second line follows from plugging in the solution for the price, (13), into the expression for the owner's value function along the equilibrium path, (8), and $k > j$ is a non-owner type with whom owners type i trade in equilibrium.[29] Then, the following equation expresses the difference in the reservation values of the owner type i when she posts the

[29] By Proposition 2 if k is an active non-owner, then all non-owners with type higher than hers are active as well. Thus, if $j \geq k$ the result that non-owners j are not idle in equilibrium would follow immediately.

deviation price and when she follows her equilibrium strategy:

$$\Delta V(\delta_i|\delta_j) - \Delta V^*(\delta_i) = \frac{1}{r + \gamma + \lambda(q)} \left[\lambda(q) S^*(\delta_i, \delta_j) - \lambda(q^*(\delta_i, \delta_k))\eta(q^*(\delta_i, \delta_k)) S^*(\delta_i, \delta_k)\right] \tag{63}$$

The left hand side of (63) is just another way of expressing $V_1(\delta_i|\delta_j) - V_1^*(\delta_i)$ and thus captures the net benefit from deviating. On the right-hand side of the equation we see the decomposition of this net benefit. It is a weighted average of two terms: (i) the payoff from deviating and (ii) the opportunity cost. The benefit from deviating is that the owner has the opportunity to trade with non-owner j. If this happens, she receives all of the surplus, so her gain is $S^*(\delta_i, \delta_j)$. The rate with which she receives an order is $\lambda(q)$, so her expected gain is $\lambda(q)S^*(\delta_i, \delta_j)$. The opportunity cost of deviating is missing out on the search option along the equilibrium path, which is the second term. In equilibrium, the owner trades with a buyer type k and receives a fraction $\eta(q^*(\delta_i, \delta_k))$ of the surplus. This order arrives at the rate $\lambda(q^*(\delta_i, \delta_k))$. Thus, while the owner i is waiting for an order from an investor type j, she is missing out on a potential order from her equilibrium trading counterparty type k. As a consequence, the owner type i would have an incentive to deviate when the fraction of the surplus she receives along the equilibrium path is too low, or if deviating provides a quick trading opportunity. Observe that given the out-of-equilibrium beliefs, the deviant expects a queue length $q = \infty$, and so the order arrival rate would depend on the behavior of the matching function in the limit. If it satisfies the Inada-type condition $\lim_{q\to\infty} \lambda(q) = \infty$, then the net benefit from deviating converges to $S^*(\delta_i, \delta_j) > 0$. Since this is a strictly profitable deviation it must be that non-owners type j are not idle in equilibrium. Since this is true for all types $i < j$, it follows that no non-owner type $j \geq 2$ is idle in equilibrium.

Similarly, we can investigate the conditions under which a given owner type j can offer favourable enough terms of trade to non-owners type $k > j$, so that the owner is not idle in equilibrium. To this end, consider an idle owner who contemplates a deviation by posting a price that is aimed at attracting some type k. The best chance she has of inducing a positive queue length is if she posts a price that leaves all of the surplus to the buyer. Given such a deviation, a non-owner type k who participates receives a utility given by

$$V_0(\delta_k|\delta_j) = \frac{\gamma V_0^*}{r+\gamma} + \frac{\lambda(q)/q}{r+\gamma} S(\delta_j,\delta_k), \tag{64}$$

where q is the induced queue length. Analogously to the case when investor j was an idle non-owner, the surplus from the trade is given by

$$S(\delta_j,\delta_k) = \frac{r+\gamma}{r+\gamma+\lambda(q)/q}\left[\Delta V^*(\delta_k) - \Delta V^*(\delta_j) + \frac{\lambda(q^*(\delta_i,\delta_k))[1-\eta(q^*(\delta_i,\delta_k))]}{q^*(\delta_i,\delta_k)(r+\gamma)} S^*(\delta_i,\delta_k)\right], \tag{65}$$

where $i < j$ is an owner type with whom non-owners type k trade in equilibrium.[30] Hence, the net benefit of participating in the submarket for a non-owner type k is

$$\Delta V^*(\delta_k) - \Delta V(\delta_k|\delta_j) = \frac{1}{r+\gamma+\lambda(q)/q}\left[\frac{\lambda(q)}{q} S^*(\delta_j,\delta_k) - \frac{\lambda(q^*(\delta_i,\delta_k))}{q^*(\delta_i,\delta_k)}[1-\eta(q^*(\delta_i,\delta_k))] S^*(\delta_i,\delta_k)\right]. \tag{66}$$

The left-hand side of the expression is simply the net benefit from participating in the deviation submarket, $V_0(\delta_k|\delta_j) - V_0^*(\delta_k)$. The right-hand side decomposes the net payoff: the first term is the expected payoff from participating in the submarket, and the second term is the opportunity cost of foregoing a potential order execution in the equilibrium submarket. If the matching function satisfies the Inada-type condition $\lim_{q\to 0}\lambda(q)/q = \infty$, then the net payoff from participating converges to $S^*(\delta_j,\delta_k)$, which is strictly positive. Moreover, this net payoff is strictly decreasing in the queue length. Thus, by continuity, the net payoff

[30] By Proposition 2, if i is an active owner, then so are all owners of type less than i. Thus, if $j \le i$, the result that owners of type j are not idle in equilibrium would follow immediately.

from participating is zero for some strictly positive, but possibly small queue length. The belief structure defined in equation (3) dictates that agents expect the largest queue length such that the buyer is indifferent between participating in the deviation submarket or not. Hence, the out-of-equilibrium beliefs imply that the posted deviation price will induce a positive queue length.[31] Since the payoffs for both the owner type j and the non-owner type k are continuous in the price, it follows that a deviation which leaves a small enough fraction of the surplus to the deviant would still induce a positive queue length. Since this is a strictly profitable deviation, it follows that owners type j will not be idle in equilibrium. Since this deviation is profitable for all idle owners type $j < k$, it follows that in equilibrium no owner type $j \leq I - 1$ is idle.

■

Lemma 2. *In equilibrium, for any type i, either all owners are sellers on some submarket, or all non-owners are buyers on some submarket, or both.*

Proof. Suppose to the contrary, that there exist some type j such that owners and non-owners of this type are both idle in equilibrium. Given proposition 3, it must be the case that $1 < j < I$. Then, consider a market (δ_i, δ_k) which opens in equilibrium with $1 \leq i < j < k \leq I$. Then, the net benefit of a seller type i who considers deviating and posting a price that attracts buyers type j, $\Delta V(\delta_i|\delta_j) - \Delta V^*(\delta_i)$, is given by equation (63). Since this is not a profitable deviation, it must be the case that the net gain is non-positive. Furthermore, since the deviation induces a queue length $q = \infty$ and the deviation payoff is strictly increasing

[31] It is worth pointing out that we are considering a deviation by a single idle owner. Thus, the ratio of buyers to sellers in the submarket in question is not well-defined: the measure of sellers on the submarket is zero. However, the queue length is a well-defined object because off the equilibrium path it is determined by the beliefs structure alone and not by the ratio of buyers to sellers.

in q, at $q = q^*(\delta_i, \delta_k) < \infty$ it follows that

$$\frac{1}{r+\gamma+\lambda(q^*(\delta_i,\delta_k))}\left[\lambda(q^*(\delta_i,\delta_k))S^*(\delta_i,\delta_j) - \lambda(q^*(\delta_i,\delta_k))\eta(q^*(\delta_i,\delta_k))S^*(\delta_i,\delta_k)\right] < 0$$
$$\Rightarrow S^*(\delta_i,\delta_j) < \eta(q^*(\delta_i,\delta_k))S^*(\delta_i,\delta_k). \quad (67)$$

Similarly, owners of type j are idle in equilibrium which implies that they cannot offer favorable enough terms of trade to non-owners type k. Hence, posting a deviation price which gives non-owners of type k the whole surplus leaves non-owners of type j with a non-positive net benefit, $\Delta V^*(\delta_k) - \Delta V(\delta_k|\delta_j)$, for any positive queue length. Since $q^*(\delta_i, \delta_k) > 0$, equation (66) implies that

$$S^*(\delta_j,\delta_k) < [1-\eta(q^*(\delta_i,\delta_k))]S^*(\delta_i,\delta_k). \quad (68)$$

But, then equations (67) and (68) imply that

$$S^*(\delta_i,\delta_k) = S^*(\delta_i,\delta_j) + S^*(\delta_j,\delta_k) < \eta(q^*(\delta_i,\delta_k))S^*(\delta_i,\delta_k) + [1-\eta(q^*(\delta_i,\delta_k))]S^*(\delta_i,\delta_k) = S^*(\delta_i,\delta_k), \quad (69)$$

which is clearly a contradiction.

■

Lemma 3. *Suppose the matching function satisfies* $\lim_{q\to\infty}\lambda(q) = \lim_{q\to 0}\lambda(q)/q = \infty$ *and that* $I > 2$. *Then, PAM cannot be an equilibrium. Moreover, the only NAM equilibrium is a single chain which starts at owner type* 1 *and ends at non-owner type* I. *That is, all possible submarkets of the type* (δ_i, δ_{i+1}) *open and no other submarket opens.*

Proof. Let us start with the PAM statement. Consider first the case that $I = 3$. Then PAM implies that the only submarket which opens is (δ_1, δ_3). But then owner type 2 is an idle seller and this is a contradiction. Next, suppose that there are at least 4 types.

By Proposition 4 owner types 1 and 2 are active sellers and non-owner types I and $I-1$ are active buyers. Suppose to the contrary that the equilibrium assignment is PAM. Then, submarkets (δ_1, δ_I) and (δ_2, δ_{I-1}) must open. But then, it follows that non-owner type 2 is an idle non-owner which is a contradiction.

We continue with the NAM statement. First, in any NAM equilibrium a submarket (δ_i, δ_j) with $j \neq i+1$ never opens. For $j \leq i$ the result is immediate from proposition 3. Next suppose that some submarket (δ_i, δ_j) opens with $j > i+1$. By Proposition 4 and the definition of NAM, it must be the case that each owner type, except I, sells to exactly one non-owner type and each non-owner type, except 1, buys from exactly one owner type. Hence, there are exactly $i-1$ submarkets where owners of type less than i participate. At the same time, it must be the case that there are exactly $j-2 > i-1$ non-owner types with valuations less than j that participate in these submarkets. But this is a contradiction.

Next, we need to show that all markets of the type (δ_i, δ_{i+1}) open. Suppose not, then there exist some i such that the above submarket does not open. But we know that no owner type $j \neq i$ will sell to a non-owner type $i+1$ by the first part of the proof. Thus, $i+1$ must be an idle buyer which is a contradiction. ■

Proposition 5. *If the matching function is linear in either the mass of sellers or the mass of buyers, intermediation cannot occur in equilibrium.*

Proof. When matching is linear in the mass of sellers, the maximization problem in (12) reduces to

$$\max_{q,j} \left\{ \mu S(\delta_i, \delta_j) - q\bar{k} \right\}, \tag{70}$$

where $\bar{k} \equiv (r+\gamma)V_0^*(\delta_j) - \gamma V_0^*$. Substituting the constraint for the buyer, $\frac{\mu}{q}[\Delta V^*(\delta_j) - p] = \bar{k}$, into the seller's value function, and differentiating with respect to q yields $\partial V_1(\delta_i)/\partial q =$

$-\bar{k}/(r+\gamma+\mu)$. Hence, $\partial S(\delta_i,\delta_j)/\partial q = \bar{k}/(r+\gamma+\mu)$. Thus,

$$\frac{\partial[\mu S(\delta_i,\delta_j)-\bar{k}]}{\partial q} = -\frac{r+\gamma}{r+\gamma+\mu}\bar{k}, \tag{71}$$

which implies that the seller wants to set the lowest possible queue length provided that $\bar{k} > 0$. Thus, sellers have an incentive to increase the price even though that drives some buyers out of the submarket. This incentive persists until the price is so high the seller extracts all of the surplus. As a consequence, the buyers outside option is $\bar{k} = 0$ in equilibrium. Thus, it is straightforward to see that owners of type $i < I$ strictly prefer to trade with type I since $\mu S(\delta_i,\delta_I) > \mu S(\delta_i,\delta_j)$ for any $j < I$. Hence, the equilibrium market structure is submarkets (i, I), $i < I$ open but no other submarket does.

When the matching function is linear in the mass of buyers, the logic is analogous similar. The seller's problem reduces to

$$\max_{q,j}\left\{\mu q S(\delta_i,\delta_j) - q\bar{k}\right\}. \tag{72}$$

Again substituting the buyer's constraint into the seller's value function and differentiating with respect to q yields $\partial S(\delta_i,\delta_j)/\partial q = -[\mu S(\delta_i,\delta_j)-\bar{k}]/(r+\gamma+\mu q)$. Thus,

$$\frac{\partial[\mu q S(\delta_i,\delta_j)-q\bar{k}]}{\partial q} = -\frac{r+\gamma}{r+\gamma+\mu q}[\mu S(\delta_i,\delta_j)-\bar{k}], \tag{73}$$

which is strictly positive as long as the seller extracts any of the gains from trade. Hence, sellers have an incentive to lower prices and attract a higher queue length. This competition for buyers pushes sellers to decrease prices until buyers extract all of the surplus in equilibrium. Thus, it is easy to see that all buyer types $j > 1$ will queue up to buy from sellers type 1 since $\mu S(\delta_1,\delta_j) > \mu S(\delta_i,\delta_j)$, for any $i > 1$. As a result, the equilibrium market structure is only submarkets $(1, j)$, with $j > 1$ open. ■

C Solution Procedure for Analytical Examples

Linear matching technology. In the case of a linear matching technology the solution procedure follows very closely the proof of Proposition 5. We reproduce it here for convenience.

When matching is linear in the mass of sellers, the maximization problem in (12) reduces to

$$\max_{q,j} \left\{ \mu S(\delta_i, \delta_j) - q\bar{k} \right\}, \tag{74}$$

where $\bar{k} \equiv (r+\gamma)V_0^*(\delta_j) - \gamma V_0^*$. Substituting the constraint for the buyer, $\frac{\mu}{q}[\Delta V^*(\delta_j) - p] = \bar{k}$, into the seller's value function, and differentiating with respect to q yields $\partial V_1(\delta_i)/\partial q = -\bar{k}/(r+\gamma+\mu)$. Hence, $\partial S(\delta_i,\delta_j)/\partial q = \bar{k}/(r+\gamma+\mu)$. Thus,

$$\frac{\partial[\mu S(\delta_i, \delta_j) - \bar{k}]}{\partial q} = -\frac{r+\gamma}{r+\gamma+\mu}\bar{k}, \tag{75}$$

which implies that the seller wants to set the lowest possible queue length provided that $\bar{k} > 0$. Thus, sellers have an incentive to increase the price even though that drives some buyers out of the submarket. This incentive persists until the price is so high the seller extracts all of the surplus. As a consequence, the buyers outside option is $\bar{k} = 0$ in equilibrium. Thus, it is straightforward to see that both owners of type 1 and type 2 strictly prefer to trade with type 3 since $\mu S(\delta_i, \delta_3) > \mu S(\delta_i, \delta_2)$. Hence, the equilibrium market structure is submarkets $(1, 3)$ and $(2, 3)$ open but submarket $(1, 2)$ does not.

When the matching function is linear in the mass of buyers, the solution procedure is analogous. The seller's problem reduces to

$$\max_{q,j} \left\{ \mu q S(\delta_i, \delta_j) - q\bar{k} \right\}. \tag{76}$$

Again substituting the buyer's constraint into the seller's value function and differentiating with respect to q yields $\partial S(\delta_i, \delta_j)/\partial q = -[\mu S(\delta_i, \delta_j) - \bar{k}]/(r + \gamma + \mu q)$. Thus,

$$\frac{\partial[\mu q S(\delta_i, \delta_j) - q\bar{k}]}{\partial q} = -\frac{r + \gamma}{r + \gamma + \mu q}[\mu S(\delta_i, \delta_j) - \bar{k}], \tag{77}$$

which is strictly positive as long as the seller extracts any of the gains from trade. Hence, sellers have an incentive to lower prices and attract a higher queue length. This competition for buyers pushes sellers to decrease prices until buyers extract all of the surplus in equilibrium. Thus, it is easy to see that both buyers type 2 and 3 will queue up to buy from sellers type 1 since $\mu S(\delta_1, \delta_j) > \mu S(\delta_2, \delta_j)$. As a result, the equilibrium market structure is only submarkets $(1, 2)$ and $(1, 3)$ open.

Leontief Matching Technology. Let us first derive the equilibrium relationship between prices and queue lengths. Observe that the Leontief matching technology coincides with the linear matching technology $m(o, n) = \mu n$ when $q \leq 1$ and with $m(o, n) = \mu o$ when $q \geq 1$. Thus, from our preceding analysis it follows that the if seller's payoff is positive, it is maximized at $q = 1$ when $\bar{k} > 0$ and at $q \geq 1$ when $\bar{k} = 0$. If competition among sellers is too high and and they do not receive any of the surplus, sellers are indifferent between any $q \leq 1$. As a result, if the equilibrium features $q^*(\delta_i, \delta_j) < 1$, it must be the case that it also features $p^*(\delta_i, \delta_j) = \Delta V^*(\delta_i)$ and if $q^*(\delta_i, \delta_j) > 1$, then $p^*(\delta_i, \delta_j) = \Delta V^*(\delta_j)$.

Market structure $(1, 3) + (2, 3)$. Next, consider the case when only submarkets $(1, 3)$ and $(2, 3)$ open. The possible combination of queue lengths in equilibrium is (i) $q(\delta_1, \delta_3), q(\delta_2, \delta_3) \geq 1$ and (ii) $q(\delta_1, \delta_3) = 1, q(\delta_2, \delta_3) < 1$. If the queue lengths were in any other combination, the equilibrium prices would be such that owners of type 3 would not be indifferent between participating in both submarkets. Namely, the prices which can support this indifference are (i) $p(\delta_1, \delta_3) = p(\delta_2, \delta_3) = \Delta V^*(\delta_3)$ and (ii) $p(\delta_1, \delta_3) = p(\delta_2, \delta_3) = \Delta V^*(\delta_2)$. When the queue lengths are such that (i) is true, the laws of motion for the masses of owners and non-owners evaluated at the steady state imply that

$$o(\delta_1) = o(\delta_2) = \frac{A}{3(1+\frac{\mu}{\gamma})}, \tag{78}$$

$$n(\delta_3) = \frac{1+\frac{\mu}{\gamma} - A(1+3\frac{\mu}{\gamma})}{3(1+\frac{\mu}{\gamma})}. \tag{79}$$

and

$$q(\delta_1,\delta_3) = \frac{\tau}{1-\tau} q(\delta_2,\delta_3) = \tau\left[\frac{1+\frac{\mu}{\gamma}}{A} - \left(1+3\frac{\mu}{\gamma}\right)\right], \tag{80}$$

where τ is the fraction of type 3 non-owners that queue up on submarket $(1,3)$ in equilibrium. Thus, the inequalities $q(\delta_1,\delta_3), q(\delta_2,\delta_3) \geq 1$ jointly require that $A \leq 1/3$, which is more restrictive than the conditions for $\tau \in (0,1)$.

When the queue lengths are such that (ii) holds, the laws of motion at steady state together with $q(\delta_1,\delta_3) = 1$ imply that $o(\delta_1) = \frac{A}{3(1+\frac{\mu}{\gamma})}$, $o(\delta_2) = \frac{A(1+3\frac{\mu}{\gamma})-\frac{\mu}{\gamma}}{3(1+\frac{\mu}{\gamma})}$, $n(\delta_3) = \frac{1-A}{3(1+\frac{\mu}{\gamma})}$, and $\tau = \frac{A}{1-A}$. Thus, $\tau \in (0,1)$ requires that $A < 1/2$ and $q(\delta_2,\delta_3) = \frac{1-2A}{A(1+3\frac{\mu}{\gamma})-\frac{\mu}{\gamma}} < 1$ requires that $A > 1/3$. Hence, the market structure can be supported as an equilibrium for $A \in (0,1/2)$.

It is clear from the preceding discussion that when $A \in (0,1/2)$ the prices and queue lengths on markets $(1,3)$ and $(2,3)$ are optimally chosen and consistent with the laws of motion. Thus, if owners type 1 do not strictly prefer to trade with non-owners type 2, then this market structure will indeed be an equilibrium. But this has to be the case because under both (i) and (ii) sellers type 1 trade at speeds μ and prices weakly higher than $\Delta V^*(\delta_2)$, whereas if a seller were to deviate and attract a positive queue length of type 2 non-owners it must be at speeds weakly less than μ and prices weakly less than $\Delta V^*(\delta_2)$.

Market structure $(1,2)+(1,3)$. When only submarkets $(1,2)$ and $(1,3)$ open the possible combinations of queue lengths consistent with seller type 1 being indifferent between participating in both markets are (i) $q(\delta_1,\delta_2), q(\delta_1,\delta_3) \leq 1$ and (ii) $q(\delta_1,\delta_2) > 1, q(\delta_1,\delta_3) = 1$.

This indifference can be supported when prices are (i) $p(\delta_1,\delta_2) = p(\delta_1,\delta_3) = \Delta V^*(\delta_1)$ and (ii) $p(\delta_1,\delta_2) = p(\delta_1,\delta_3) = \Delta V^*(\delta_2)$. When both queue lengths are less than 1 the laws of motion at steady state imply that

$$n(\delta_2) = n(\delta_3) = \frac{1-A}{3(1+\frac{\mu}{\gamma})}, \tag{81}$$

$$o(\delta_1) = \frac{A(1+3\frac{\mu}{\gamma}) - 2\frac{\mu}{\gamma}}{3(1+\frac{\mu}{\gamma})}. \tag{82}$$

As a result, the queue lengths are given by

$$q(\delta_1,\delta_2) = \frac{1-\phi}{\phi} q(\delta_1,\delta_3) = \frac{1-A}{\phi\left[A(1+3\frac{\mu}{\gamma}) - 2\frac{\mu}{\gamma}\right]}, \tag{83}$$

where ϕ is the fraction of type 1 owners who participate on submarket $(1,2)$. Thus, the restriction that both queue lengths are less than unity requires $A \geq 2/3$, which also ensures $\phi \in (0,1)$.

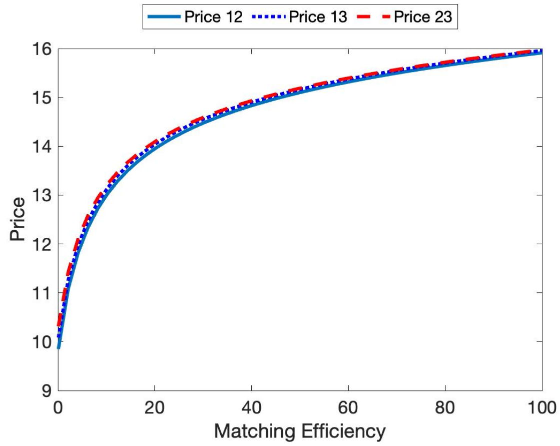
When $q(\delta_1,\delta_2) > 1$, the laws of motion together with $q(\delta_1,\delta_3) = 1$ lead to $o(\delta_1) = \frac{A}{3(1+\frac{\mu}{\gamma})}$, $n(\delta_2) = \frac{1+2\frac{\mu}{\gamma} - A(1+3\frac{\mu}{\gamma})}{3(1+\frac{\mu}{\gamma})}$, $n(\delta_3) = \frac{1-A}{3(1+\frac{\mu}{\gamma})}$, and $\phi = \frac{2A-1}{A}$. As a result, $q(\delta_1,\delta_2) > 1$ and $\phi > 0$ require that $A \in (1/2, 1/3)$. Thus, the market structure can be supported as an equilibrium for $A \in (1/2, 1)$.

Analogously to the previous market structure we have examined, the only potentially profitable deviation is for sellers type 2 to post prices and queue lengths that attract type 3 non-owners. But such a deviation cannot be profitable. In both (i) and (ii) non-owners type 3 receive the asset at a rate μ at prices weakly less than $\Delta V^*(\delta_2)$. Thus, a deviant would never be able to extract positive surplus from trade with a type 3 non-owner.

Market structure All. When all markets open the only possible combination of the queue lengths consistent with the indifference condition for owners type 1 and non-owners type 3 is $q(\delta_1,\delta_2) \geq 1, q(\delta_1,\delta_3) = 1, q(\delta_2,\delta_3) \leq 1$. These indifference conditions together with the optimal price setting behavior of sellers implies that prices on all submarkets are equal to the reservation value of type 2 investors. The laws of motion at steady state together with the restriction on $q(\delta_1, \delta_3)$ imply that $o(\delta_1) = \frac{A}{3(1+\frac{\mu}{\gamma})}$, $o(\delta_2) = \frac{A(1+3\frac{\mu}{\gamma})-\frac{\mu}{\gamma}}{3(1+\frac{\mu}{\gamma})}$, $n(\delta_3) = \frac{1-A}{3(1+\frac{\mu}{\gamma})}$, and $\tau = \frac{A}{1-A}\phi$. Thus, the restrictions on the queue lengths transform to $\frac{1+\frac{\mu}{\gamma}}{2+3\frac{\mu}{\gamma}+\phi} \leq A \leq \frac{1+2\frac{\mu}{\gamma}}{2+3\frac{\mu}{\gamma}-\phi}$. Noting that τ, ϕ can take any values inside $(0, 1)$, it is easy to verify that that these restrictions are consistent with $A \in (1/3, 2/3)$. Lastly, we note that by construction queue lengths and prices on all submarkets are optimally chosen and buyers and seller are indifferent between participating in the different submarkets. Hence, this proposed market structure is an equilibrium for $A \in (1/3, 2/3)$.

Market structure $(1, 2)+(2, 3)$. It is easy to see that when only submarkets $(1, 2)$ and $(2, 3)$ open there aren't any indifference conditions that need to be satisfied and so this market structure is consistent with any level of the asset, A. The masses of owners and non-owners at steady state do depend on the level of asset supply, however. It is straightforward to verify that when $A \leq \frac{1+\frac{\mu}{\gamma}}{2+3\frac{\mu}{\gamma}}$ the masses of agents are such that $o(\delta_1) \leq n(\delta_2)$ and $o(\delta_2) \leq n(\delta_3)$; when $\frac{1+\frac{\mu}{\gamma}}{2+3\frac{\mu}{\gamma}} \leq A \leq \frac{1+2\frac{\mu}{\gamma}}{2+3\frac{\mu}{\gamma}}$, $o(\delta_1) \leq n(\delta_2)$ and $o(\delta_2) \geq n(\delta_3)$; when $A \geq \frac{1+2\frac{\mu}{\gamma}}{2+3\frac{\mu}{\gamma}}$, $o(\delta_1) \geq n(\delta_2)$ and $o(\delta_2) \geq n(\delta_3)$. As a result the steady state masses for owners of types 2 and 3 are $o(\delta_2) = \frac{A(1+2\frac{\mu}{\gamma})}{3(1+\frac{\mu}{\gamma})^2}$, $o(\delta_3) = \frac{A}{3} + \frac{\mu}{\gamma}o(\delta_2)$; $o(\delta_2) = \frac{A(1+3\frac{\mu}{\gamma})-\frac{\mu}{\gamma}}{3(1+\frac{\mu}{\gamma})}$, $o(\delta_3) = \frac{A+\frac{\mu}{\gamma}}{3(1+\frac{\mu}{\gamma})}$; $o(\delta_2) = \frac{(1+\frac{\mu}{\gamma})^2-(1-A)(1+2\frac{\mu}{\gamma})}{3(1+\frac{\mu}{\gamma})^2}$, $o(\delta_3) = \frac{A+\frac{\mu}{\gamma}}{3(1+\frac{\mu}{\gamma})}$, respectively.

The net present value of welfare at steady state is given by $\int_0^\infty e^{-rt}[o(\delta_2)\delta + o(\delta_3)]dt = \frac{o(\delta_2)\delta+o(\delta_3)}{r}$. Given this and the steady state masses we have derived above for each market structure, straightforward algebra leads to the inequalities in table 1.



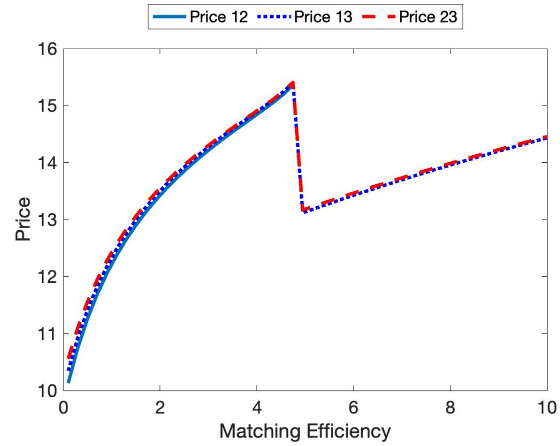


Figure 3: Prices in Random (Left) and Competitive (Right) Search.
Prices charged by type i seller to type j buyer for different levels of matching efficiency.

D Numerical Results Supplement

We begin the numerical results supplement with a three-type version of the competitive and random search models. Investors can have the following asset valuations: $\delta_1 = 0$, $\delta_2 = 1/2$, and $\delta_3 = 1$. Types 1 and 3 are natural customers, whereas type 2 may act as an intermediary. The discount rate is set at 5% and the arrival rate of preference shocks is $\gamma = 1$, following Hugonnier et al. (2014). The asset supply is $A = 0.5$, unless noted otherwise, and the matching efficiency is $\mu = 1$. The underlying distribution of utility types is assumed to be uniform and we assume equal bargaining weights for buyers and sellers in the random search model. Finally, an important difference between our model and the random search case is that meeting rates are endogenous and formed through a matching function. To make the comparison as transparent as possible, we choose the telephone line matching function, $m(o,n) = \mu\frac{on}{o+n}$, which implies that the meetings rate in our model have exactly the same functional form as in the random search model.

In Figure 3, we plot the prices for various seller-buyer couples in the two models as matching efficiency grows, while the asset supply is equal to 0.3. The asset quantity pins down a specific investor type as the "marginal type" in the frictionless benchmark. As μ

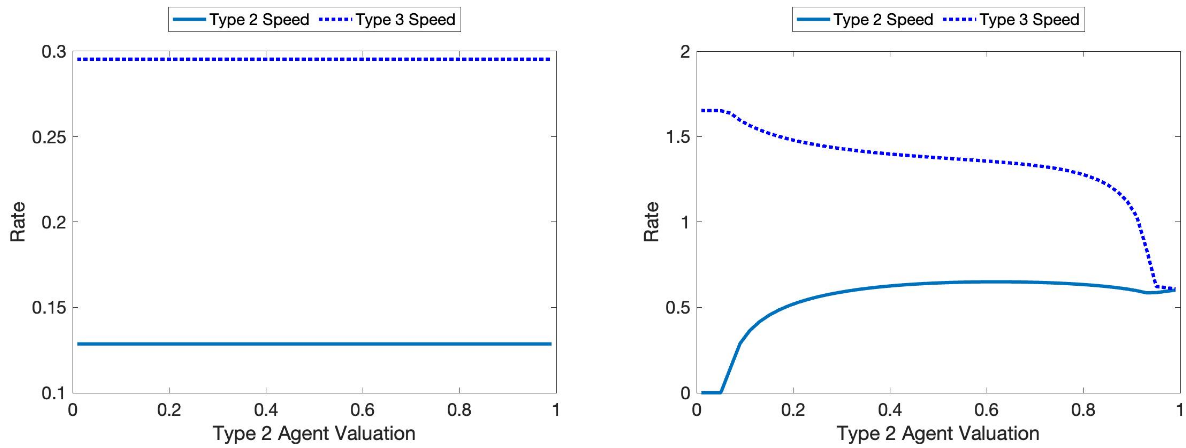


Figure 4: Speed of Trade in Random (Left) and Competitive (Right) Search.
Matching rates for type 2 and 3 buyers for different levels of the type 2 valuation.

increases, prices converge to the reservation utility of the marginal type (in this case, type 3), as both models converge to the frictionless case. There are two important takeaways from Figure 3: first, the convergence in the random search model is much slower than in our model. Our model delivers prices of similar magnitude for levels of matching efficiency that are an order of magnitude lower. The reason is of course that meeting rates in our model have an endogenous part which comes from the self-selection of agents in different submarkets. The quantitative impact of this endogenous part on prices can be substantial. Second, there is a discrete jump in prices in our model when μ is around 5, while the transition in the random search model is smooth. This jump is an implication of the endogenous market structure in our model, which changes for different levels of matching efficiency; that is, submarket (δ_1, δ_2) shuts down for large values of μ and, as a result, prices in the other two markets change discretely. In contrast, the market structure in random search is the same for all parameter values and prices change smoothly.

Figure 4 plots the speed of trade for buyers of type 2 and 3 in the two models for various levels of the flow utility of the middle type. It captures an important difference between the random and competitive search models: the levels of investor valuations do not affect trading

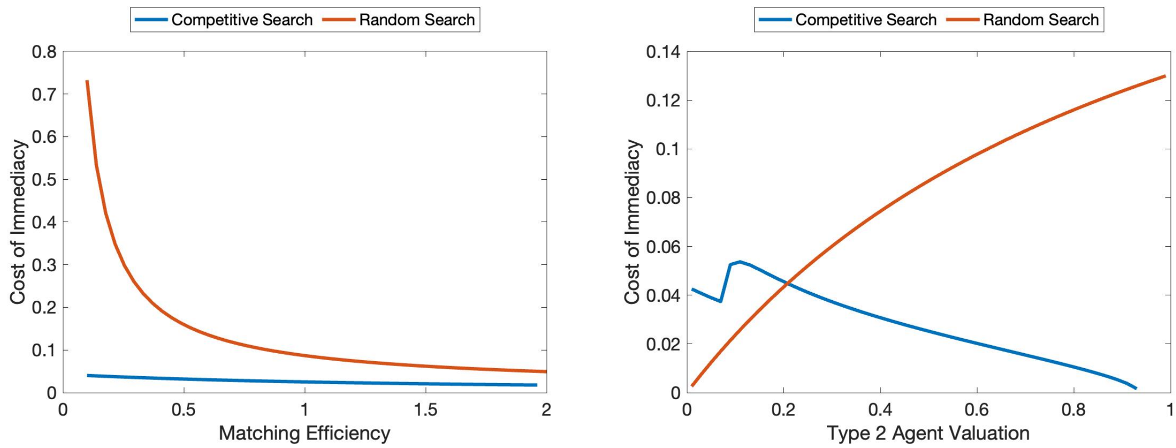


Figure 5: Cost of Immediacy in Random and Competitive Search.
Cost of immediacy for different levels of matching efficiency (left) and type 2 valuation (right) in the random (red line) and competitive (blue line) search models.

speed under random search, while they have a large impact under competitive search. Under random search, investor valuations affect only asset prices but trading speeds are fixed. On the other hand, under competitive search, trading speeds are endogenous and reflected in prices, because investors choose counterparties and the market structure is endogenous. As the asset becomes more valuable for middle-type agents, they keep the asset longer (the speed of trade for type 3 decreases) and they buy the asset faster from low-type investors 1.

The differences in how investor valuations affect trading speeds in the two models are reflected in their implications for the cost of immediacy in Figure 5. Our measure of the cost of immediacy captures how much more expensive is for a customer of type 3 to buy the asset quickly from the middle-value agent 2 (who plays the role of the intermediary) versus buying it at a lower speed from the low-value investor 1. More concretely, the cost if immediacy is defined as the percentage change of price $p(\delta_2, \delta_3)$ versus $p(\delta_1, \delta_3)$ over the percentage change in the expected order filling time from 2 versus 1. This elasticity decreases sharply as μ increases in the random search model, while it is relatively flat in our model. The reason is that under random search trading speeds are fixed as prices change, while in our model trading speeds adjust and make the impact of price changes on the cost of

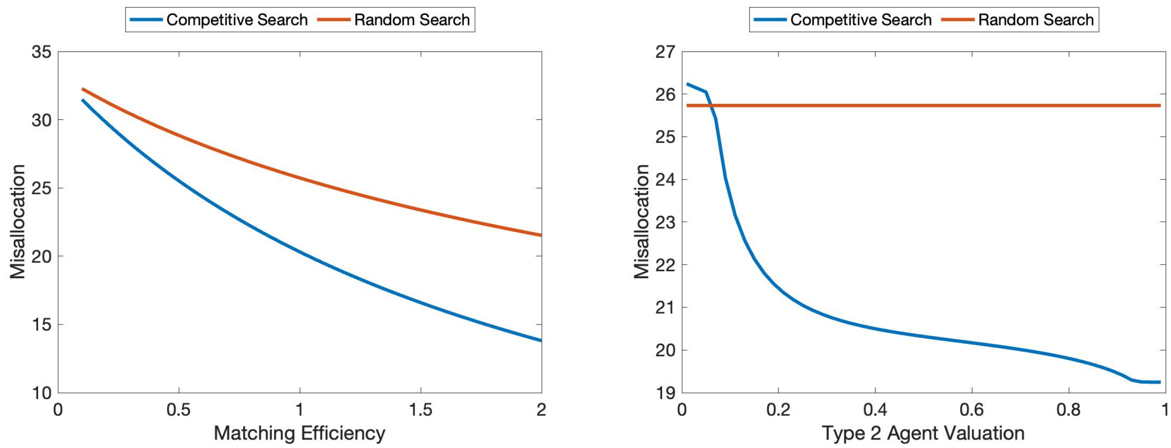


Figure 6: Misallocation in Random and Competitive Search.
Misallocation for different levels of matching efficiency (left) and type 2 valuation (right) in the random (red line) and competitive (blue line) search models.

immediacy smaller. The response of trading speeds may even change the direction of cost of immediacy in some cases, as the right panel of Figure 5 shows. As δ_2 increases, the cost of immediacy increases in the random search model because agents 2 can be compensated through prices only. In our model, they are also compensated with higher trading speeds and, as a result, the cost of immediacy for customers 3 decreases.

The two models also differ in their implications for misallocation and intermediated trade volume, as seen in Figures 6 and 7 respectively.[32] We follow Hugonnier et al. (2021) and define asset misallocation as the difference in the asset holding distribution in the model equilibrium versus the frictionless benchmark. As μ increases (left panel of Figure 6), misallocation drops in both models, but the decrease is faster under competitive search. This reflects that, on top of prices, there are two more margins to adjust in our model: trading speeds and market structure. These two margins also imply that as δ_2 increases (right panel of Figure 6), misallocation sharply decreases with competitive search, while it does not change in the random search case. Finally, these mechanisms differentiate the two models in terms of intermediation predictions. Trade that goes through the middle-value agent 2

[32] The results for total trade volume are qualitatively similar to those for intermediated trade volume.

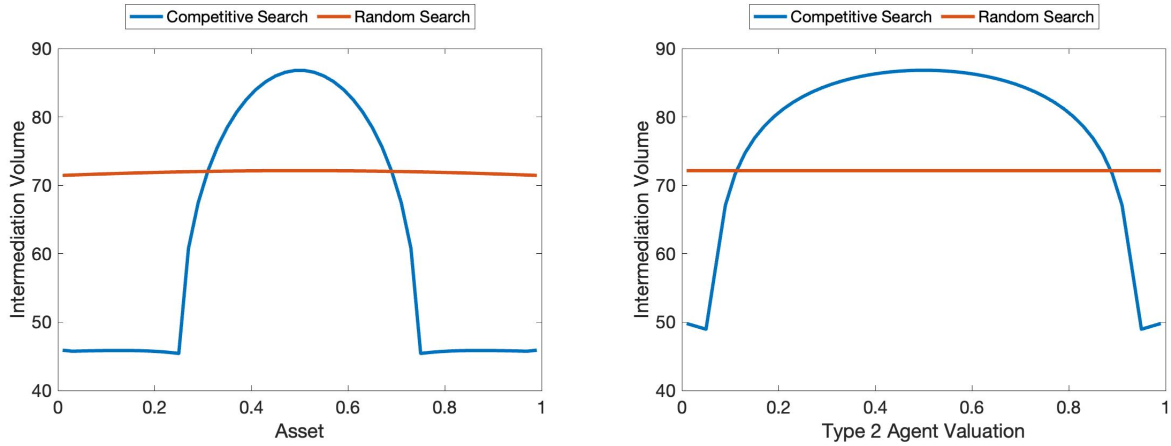


Figure 7: Intermediated Trade in Random and Competitive Search.
Fraction of trade intermediated by at least one dealer for different levels of matching efficiency (left) and type 2 valuation (right) in the random (red line labeled) and competitive (blue line) search models.

(as fraction of total trade volume) barely changes for various levels of asset supply or δ_2 valuations in the random search model. On the other hand, there are large changes in intermediated volume predicted by our model. For low and high levels of asset supply and δ_2, the scope of intermediation is limited: agents 2 act more like customers than intermediaries and the market structure features relatively few (δ_1, δ_2) and (δ_2, δ_3) trades. For intermediate values of asset supply and δ_2, intermediation becomes the key mechanism through which the market allocates the asset to the high-value customers 3. This prediction, which rests on the endogeneity of the market structure, is absent from the random search framework.

Next, we want to compare the two models regarding their implications for intermediation chains, an important feature of OTC markets, as documented by Li and Schürhoff (2019) and Hollifield et al. (2017) among others. To this end, we increase the number of types to five, implying that there can be three types of intermediation chains between customers $(1, 5)$ and intermediaries $(2, 3, 4)$. That is, an asset starting from type 1 can end up in the hands of type 5 following a chain with one, two, or three intermediation steps. One could, of course, examine longer intermediation chains, but Li and Schürhoff (2019) show that chains including up to three dealers account for over 98% of observed trades in the municipal bond market. Hence, we limit our attention to chains with one, two, or three intermediation steps. Figure 8 plots the effects of changes in the quantity of the asset on various measure associated with intermediation chains in the two models.

In the top panel, we plot the fraction of trades following the asset that go through chains of one, two, or three intermediaries. The vast majority of trades are customer-dealer-customer (CDC) trades in both models, but there are other chains especially for intermediate asset values. Importantly, in our model the changes are more abrupt, since we see large changes in the magnitude of trades going through different chains for small changes in the value of the asset. This is another implication of the fact that, under competitive search, the market structure is endogenous and changes when parameters change. For some parameter values CDC trades disappear, as various submarkets become inactive and the asset goes through multiple steps before reaching investor 5. Recall that having long intermediation chains is an efficient outcome in our model. This implies that two-step chains is the most efficient way for the asset to travel from investor 1 to investor 5 for intermediate asset levels.[33]

In the middle and bottom panel of Figure 8, we plot the fraction of trades in which each dealer type participates (middle panel) and the mark-up that each dealer type charges when she intermediates assets in CDC chains (bottom panel). We interpret the fraction of trades in which a dealer participates as a measure of the dealer's *centrality*. Similarly, the mark-up charged by each dealer (the bid-ask spread normalized by the bid price) is a measure of the cost of intermediation services provided by the dealer. The comparison of a dealer's centrality with the mark-up she charges is informative regarding the existence of a centrality premium or discount in the two models. That is, whether dealers who intermediate larger

[33] Following Hugonnier et al. (2014), we appropriately condition on the event that the agents involved in intermediation chains do not experience a preference shock between buying and selling the asset. We do so because investors switch back and forth being a dealer and customer in the model, while these roles are more stable in the data. See Section 4.2 in Hugonnier et al. (2014).

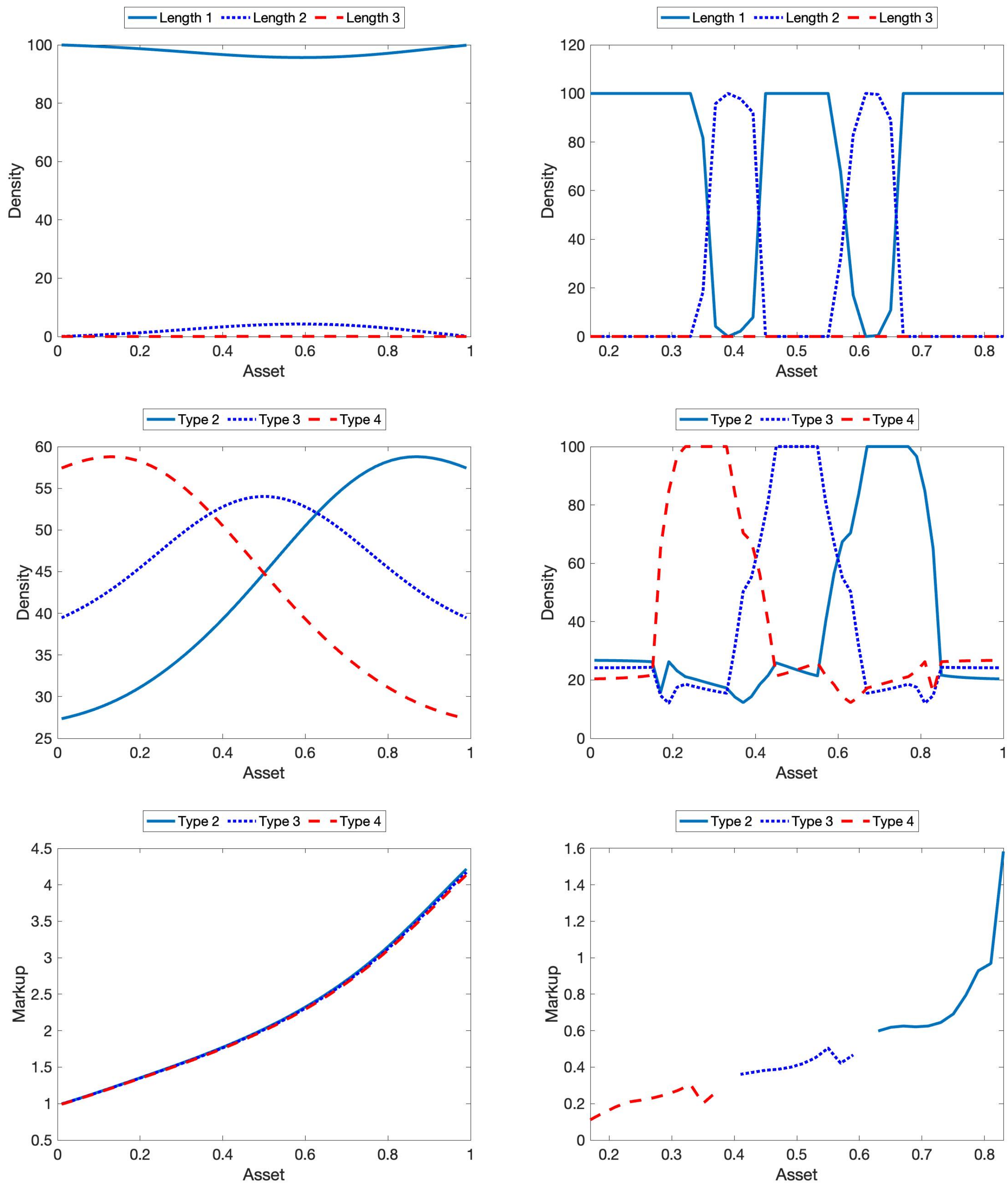


Figure 8: Intermediation Chains in Random (Left) and Competitive (Right) Search. Top Panel: fraction of trades going through chains that consist of one, two, and three intermediaries. Middle Panel: fraction of trades in which each dealer type participates. Bottom Panel: mark-up charged by each dealer type in intermediation chains. All variables are shown for different levels of asset quantity.

volumes charge higher or lower mark-ups; see Neklyudov (2019) for a concise summary of the evidence. As we see in Figure 8, neither of the two models makes clear predictions regarding the relationship between dealer centrality and mark-up.[34] In the context of the random search model, this result had also been found by Hugonnier et al. (2021). Here we show that price competition alone is not enough to generate a clear relationship between these variables. In both models, intermediation volume strongly depends on the level of the asset, but asset prices strictly follow investors' reservation values and their ranking does not change for different asset levels.

In the competitive search model, trading speed is an outcome of the endogenous meeting process between investors. Thus, it depends on the form of the matching function. So far, we have used a telephone line matching function to facilitate a clear comparison with the random search model. Next, we also look at the Cobb-Douglas case, $m(o, n) = \mu o^{1-\alpha} n^{\alpha}$, the most common matching function in the search and matching literature. Generally, the implications of the competitive search model with a Cobb-Douglas matching function are similar to the case of the telephone matching function, except that the market structure changes less abruptly. This is because Cobb-Douglas satisfies the Inada-type conditions of Proposition 4, while the telephone matching does not. However, the Cobb-Douglas case allows us to perform comparative statics with respect to the meeting elasticity, α, and we present the results for selected variables of interest.

In Figure 9, we show how the level of misallocation (left panel) and speed of trade by buyer type (right panel) varies for different levels of the meeting elasticity in a version of the model with three valuation types. Raising α increases the importance of buyers for matching and gives them a larger part of the surplus through lower prices. This incentivizes buyers 2 and 3 to acquire the asset and lowers misallocation, since these are the high-value investors.

[34] It should be noted that the markups are not well-defined for all parameter values, since the relevant submarkets do not always open. This explains why the lines on the bottom right panel of Figure 8 are graphed for only some levels of the asset supply.

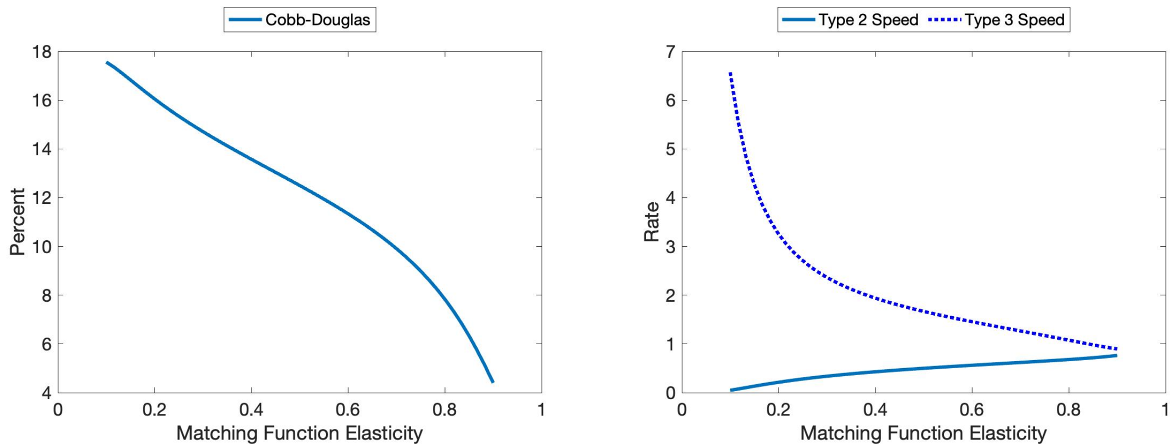


Figure 9: Misallocation (Left) and Speed of Trade (Right) in Competitive Search. Misallocation and buying speed for different levels of the meeting elasticity with a Cobb-Douglas matching function.

Importantly, since 2 intermediates the asset for type 3, they also split the gains from type's 3 higher surplus when buying from type 1. As a result, types 2 have the largest gain from the increase in α, because they benefit directly (as buyers from 1) and indirectly (as dealers for 3). They buy the asset faster and keep the asset longer. These results highlight an interesting point about intermediation in the competitive search model: middle-value agents who act as dealers benefit more than customers when buyers acquire greater importance in the matching process (the equivalent of bargaining power in the random search model).

In Figure 10, we plot the effects of changes on asset quantity on intermediation chains in the model with five investor types and Cobb-Douglas matching. We present the results for two levels of matching efficiency: on the left panel, μ is equal to one (as in all other experiments), while in the right panel we set it equal to ten. Focusing on the left panel, the results are similar to the telephone matching case (see Figure 8). The only difference is that with Cobb-Douglas the density of chains involving both one and two dealers is almost always positive, whereas for some asset levels only chains with either one or two dealers exist under telephone matching. As mentioned earlier, this is due to the fact that the market structure changes more abruptly with telephone line matching, with some markets closing or opening

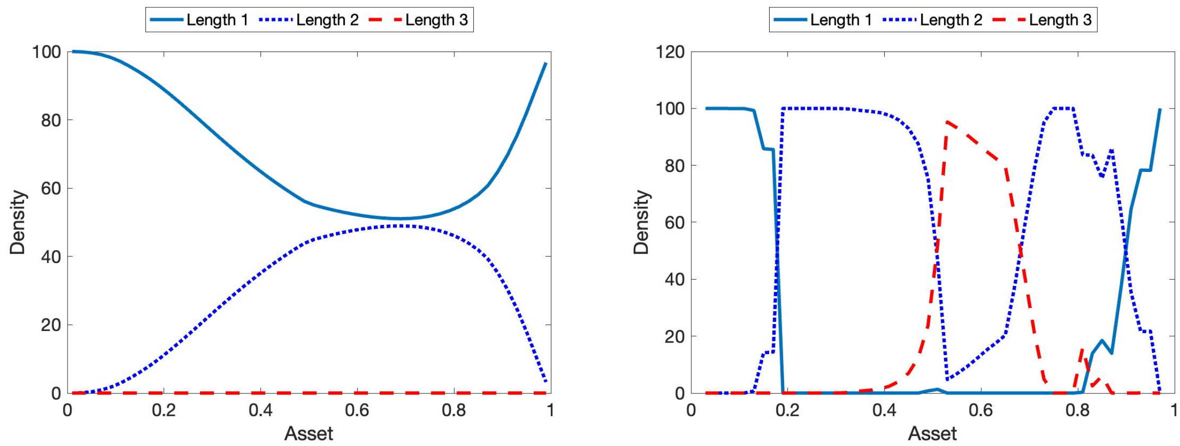


Figure 10: Intermediation Chains with Cobb-Douglas Matching.
Fraction of trades going through chains that consist of one, two, and three intermediaries for the Cobb-Douglas matching function when matching efficiency equals one (left panel) and ten (right panel).

for small changes of the asset supply. When matching is very efficient, as shown in the right panel of Figure 10, we obtain another interesting result: for intermediate asset levels, almost all intermediation chains involve either two or three dealers in the Cobb-Douglas case. Chains involving one dealer play a role in the model only for low and high levels of asset supply. This is in sharp contrast with both the telephone line matching case and the random search model with Nash bargaining (Figure 8), where CDC chains are very common. The intuition, of course, is that with high matching efficiency the intermediation costs are low, so it is efficient to form chains with many dealers.

Given the importance of intermediation for our approach and in light of our findings in Section 4 that intermediation may have limited welfare improving properties, we also compute the welfare gains from intermediation. Specifically, we compute the difference in welfare between the full model and a restricted version in which we do not allow a submarket which is essential for intermediation to open. That is, we compute the maximum welfare given an additional constraint imposing that type 1 owners cannot sell the asset to type 2 non-owners. Essentially, we shut down intermediation and compare the welfare under the extra restriction with the welfare in the full model for the telephone line and the Cobb-Douglas matching functions. Figure 11 plots the welfare gains due to intermediation in comparative statics with respect to asset quantity, the valuation of type 2, and matching efficiency.

In general, the shape of welfare gains from intermediation for different parameter values has a similar shape for both matching functions. The magnitude of the gains though is larger in the Cobb-Douglas case. Focusing on the magnitude of the gains, Cobb-Douglas satisfies the Inada-type conditions of Proposition 4 which guarantee that the extent of intermediation is maximized in the competitive search equilibrium. As a result, it features more intermediation and a larger welfare gain from intermediation than the telephone matching function. The shape of welfare gains for different parameter values is also intuitive. The welfare gains are increasing in asset quantity (top left panel) and the valuation of type 2 (top right panel), up to some threshold value after which they slightly decrease. The greater the asset supply, the greater the misallocation occurring without intermediation; hence, allowing for intermediation brings a larger welfare improvement. Similarly, when the asset is more valuable to type 2, there is a greater welfare loss when type 2 is not allowed to receive the asset from type 1. Interestingly, this relationship is non-monotonic: welfare gains slightly decrease when δ_2 becomes very close to δ_3.

Similarly, welfare gains increase with matching efficiency up to some value but they decrease when matching efficiency exceeds a threshold level. The reason for these non-monotonicities is that the restricted model comes closer to the full model for high values of δ_2 and μ. For large δ_2, type 2 non-owners become very similar to type 3 customers; for large μ, the asset reaches type 3 customers very quickly even without intermediation. As a result, the importance of intermediation for welfare declines for large levels of δ_2 and μ. It is worth noting that this comparative static has an interesting implication with some policy relevance. As markets become more efficient in connecting investors, the welfare-improving effects of intermediation may decline. Hence, the importance of intermediaries for welfare may also become smaller as price competition and high matching efficiency are enough to transfer the asset to high-value investors quickly.

In Figure 12, we plot further comparative statics for the Cobb-Douglas matching function to complement those shown in Figure 9. All comparative statics are with respect to the meeting elasticity, α. The meeting elasticity is fully reflected in asset prices (equation 13), giving more of the surplus to non-owners as α approaches one (top left panel). The top right panel plots the fractions of trade that go through chains of different lengths. The majority of trades goes through a CDC chain but the fraction of CDDC trades (with two dealers involved) is sizeable, reaching 40% for medium values of α. Finally, in the bottom panel we see the relationship between dealer centrality (as captured by the fraction of trades intermediated by each dealer type; bottom left panel) and markups (bottom right panel). Again, the model does not seem to provide a clear prediction regarding a centrality premium or discount.

In Figure 13, we plot a set of similar variables but for a different comparative static using the Cobb-Douglas matching technology: a mean-preserving spread in investor valuations. Specifically, we fix $\delta_2 = 5$ and let δ_1 vary from 4 to 0 and δ_3 from 6 to 10 in equal increments. Influenced by Lagos et al. (2011), we think of these changes in investor valuations as simulating a "crisis" in the model: periods in which investors become increasingly polarized in their asset valuations. Prices in different submarkets diverge (top left panel), reflecting the different investor valuations for the asset, in a model with three investor types. Interestingly, diverging prices do not affect the trading speeds for buyers of different types, as shown in the bottom left panel of Figure 13. Moving to the model with five investor types, the right top and bottom panels depict the fraction of trades intermediated by dealers of different types and the markups they charge, respectively. As valuations increasingly diverge, investor 3 emerges as the central intermediary, intermediating increasingly larger fractions of trades

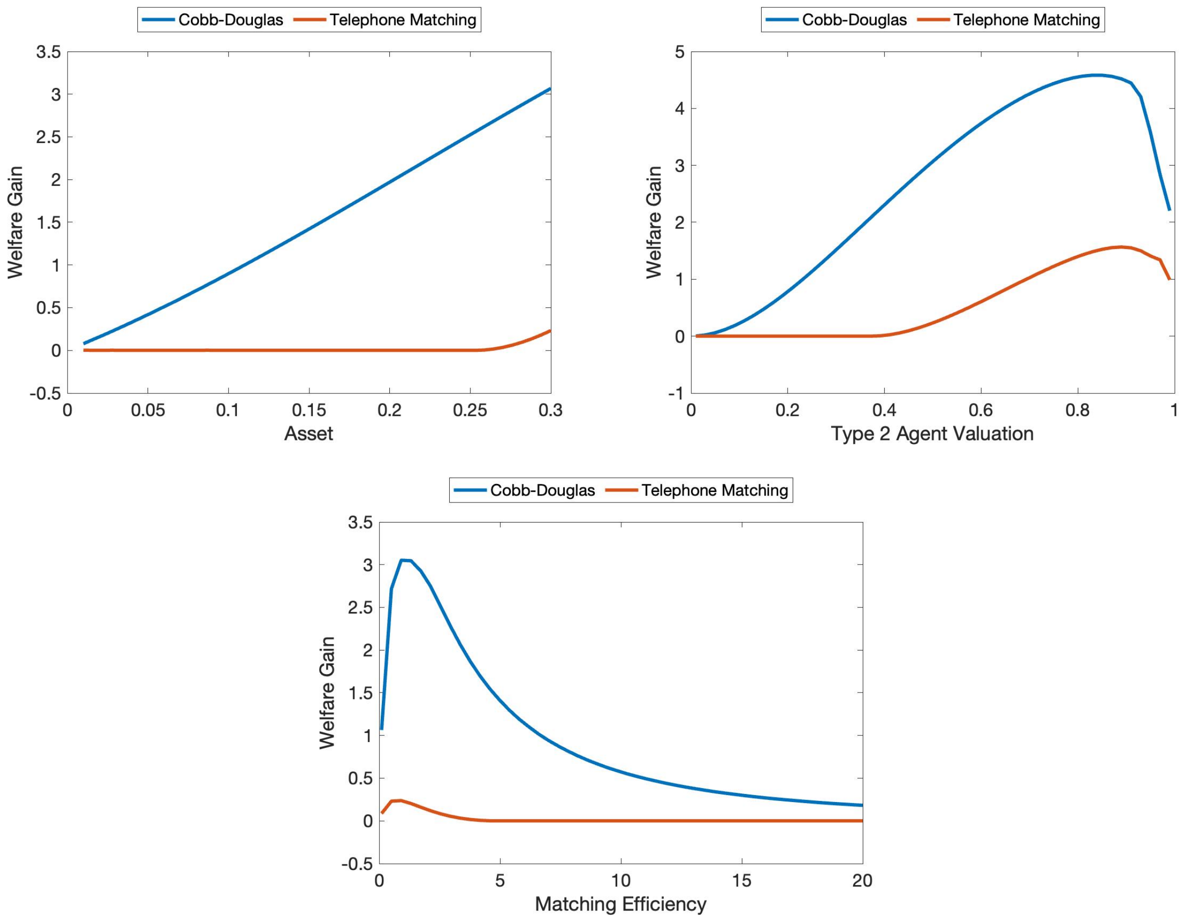


Figure 11: Welfare Gains from Intermediation in Competitive Search.
The welfare gains are the difference in aggregate welfare computed in the full model and a restricted model version for which owners of type 1 are not allowed to sell to non-owners of type 2. The models are solved under competitive search and for a Cobb-Douglas (blue line) and a telephone line (red line) matching function. The top left panel shows welfare gains from intermediation for different levels of asset quantity. The top right panel shows welfare gains from intermediation for different levels of the valuation of type 2. The bottom panel shows welfare gains from intermediation for different values of the matching efficiency.

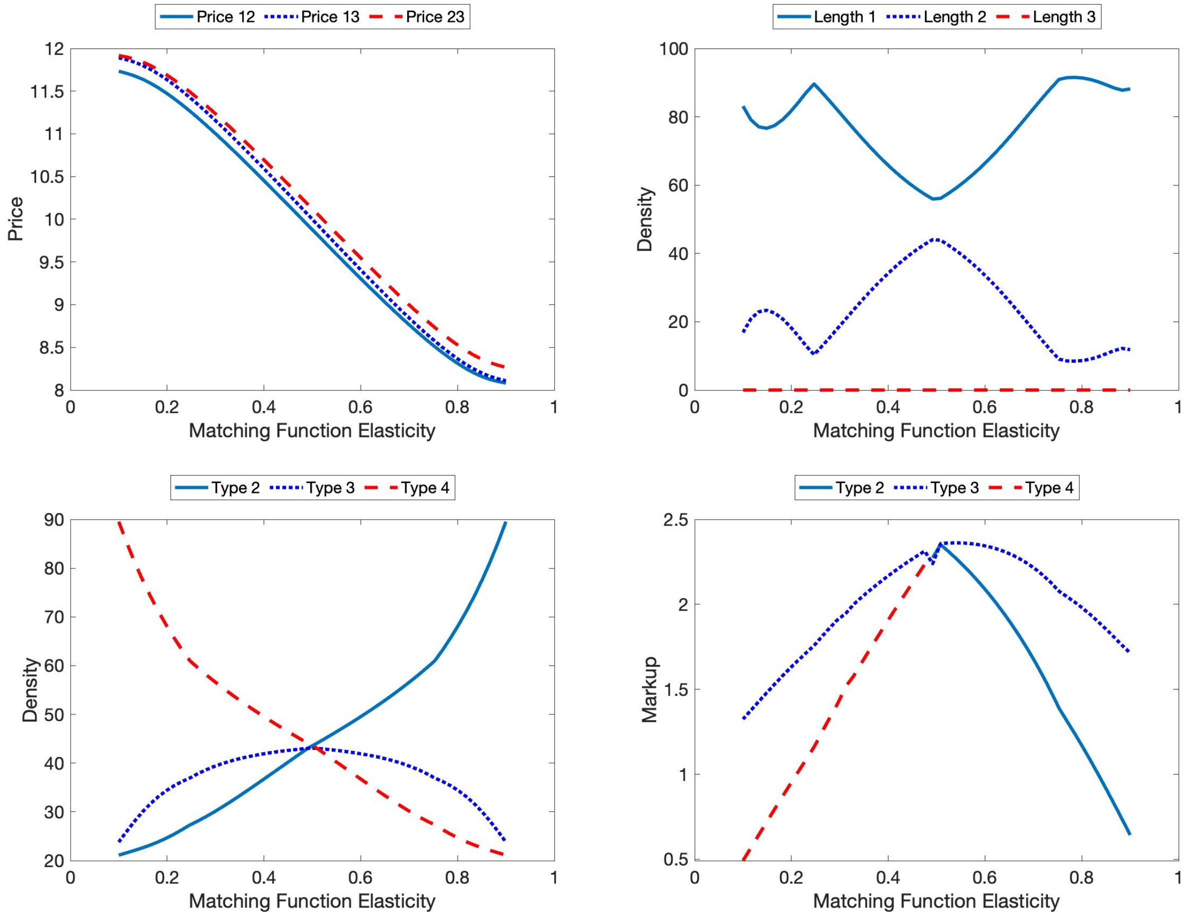


Figure 12: Comparative Statics with Cobb-Douglas Matching.
Prices (top left panel), fractions of trade going through chains of different length (top right panel), fraction of trade intermediated by different dealers (bottom left panel), and mark-ups charged by different dealers (bottom right panel) for different levels of the meeting elasticity with a Cobb-Douglas matching function.

compared to intermediaries of types 2 and 3. All intermediaries charge higher markups (a prominent feature of crises), which are strikingly similar: another instance in which the model has difficulty generating either a centrality premium or discount.

We close the Appendix by providing information on two alternative measures for the cost of immediacy in the random and competitive search models. As a reminder, the main measure we used in Section 5 captured the cost of immediacy from the perspective of type 3 buyers. Specifically, it was defined as the percentage change of price $p(\delta_2, \delta_3)$ versus $p(\delta_1, \delta_3)$

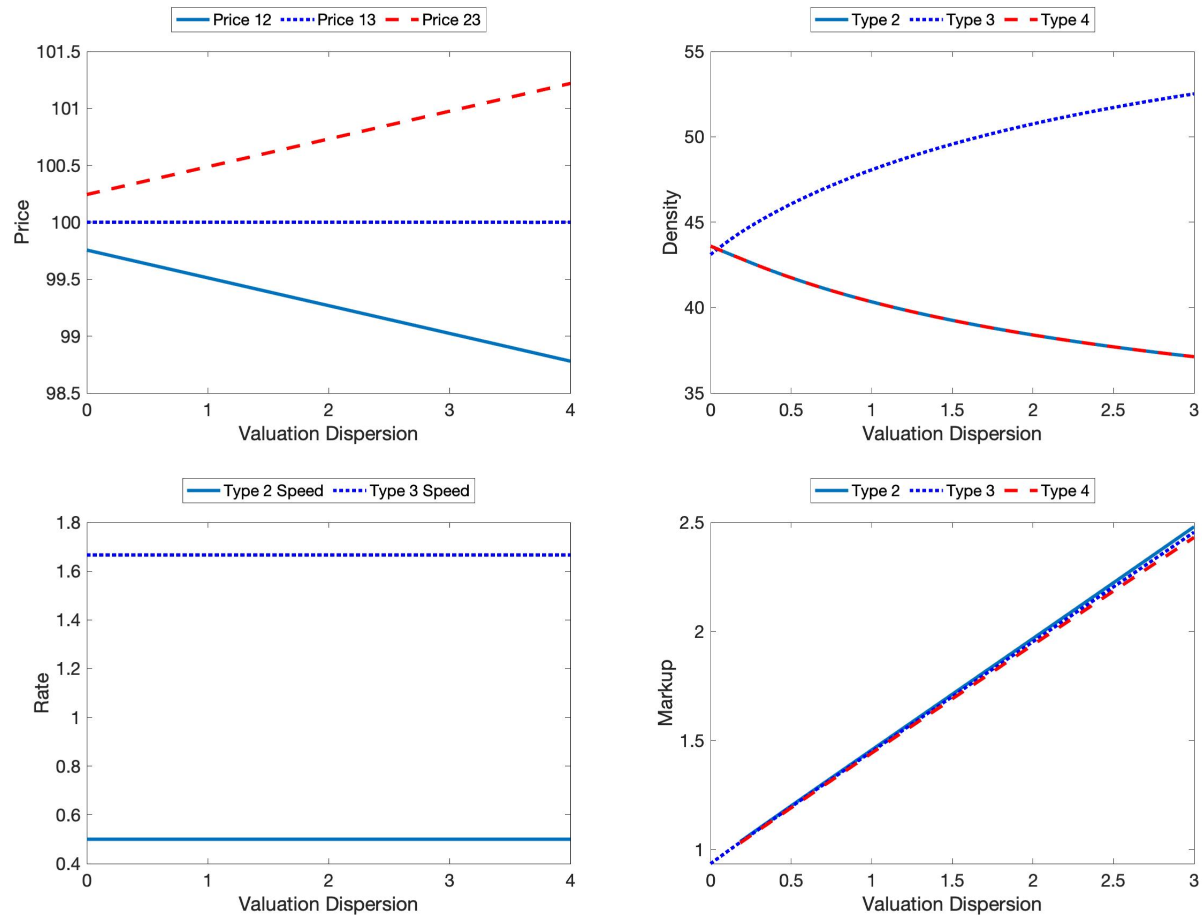


Figure 13: Mean-Preserving Spread with Cobb-Douglas Matching.
Prices (top left panel), fraction of trade intermediated by different dealers (top right panel), buying speed for different types (bottom left panel), and mark-ups charged by different dealers (bottom right panel) for mean-preserving spreads in investors' valuations with a Cobb-Douglas matching function.

over the percentage change in the expected order filling time from investor type 2 versus 1. The first additional measure we employ is the bid-ask spread charged by type 2 who plays the role of intermediary in the three-type model (the difference $p(\delta_2, \delta_3) - p(\delta_1, \delta_2)$, to be precise). The bid-ask spreads for the comparative statics in the random and competitive search model are shown in Figure 14. In the left panel of Figure 14, we see that bid-ask spreads decline more rapidly in the competitive than the random search model, when matching efficiency improves. Intuitively, dealer rents disappear faster in competitive search, because of the

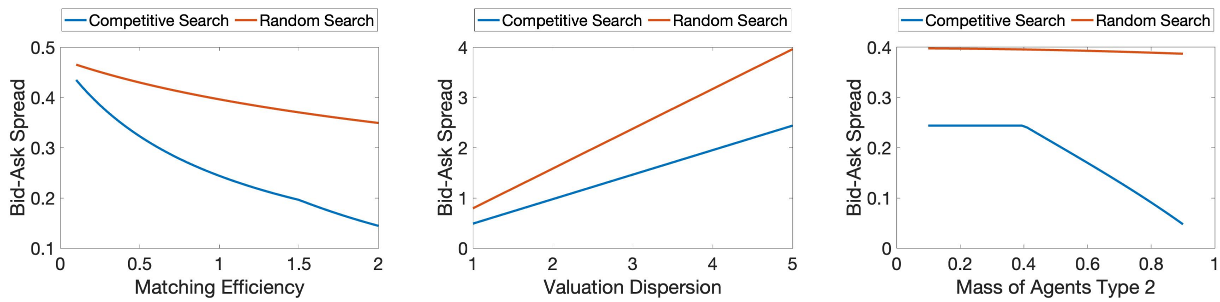


Figure 14: Bid-ask Spreads in Random and Competitive Search.
Bid-ask spreads for different levels of matching efficiency μ (left), valuation dispersion $|\delta_1 - \delta_2| = |\delta_3 - \delta_2|$ (middle), and the mass of type 2 agents $f(\delta_2)$ (right) in the random (red line) and competitive (blue line) search models.

extra margin of adjustment. On top of prices, agents self-select into different submarkets and the convergence to the frictionless outcome (with zero dealer rents) is faster than in the random search framework. Under a mean-preserving spread (middle panel), we see exactly the same behavior as in Figure 1, and the same intuition applies. Finally, when the mass of dealers increases (right panel), dealer rents deteriorate in both models, but the speed is faster in the competitive search model because of the endogenous adjustment of order filling rates. Another way to see that is to notice the kink in the bid-ask spread when the mass of dealers becomes greater than 0.4 in the competitive search model, due to the change in market structure.

The final measure we comment on is the analogue of the cost of immediacy used in the main text but for sellers of type 1 instead of buyers of type 3: the percentage change of price $p(\delta_1, \delta_2)$ versus $p(\delta_1, \delta_3)$ over the percentage change in the expected order filling time when trading with investor type 2 versus 3. A comparison between Figures 1 and 15 shows that the behavior of the two measures is identical. The reason, of course, is that they are the flip side of each other: whatever cost the type 3 buyer has to pay for immediacy is the premium that ends up in the hands of type 1 seller in order to incentivize them to participate in these trades. As a result, the behavior of this measure is the mirror image of the cost of immediacy

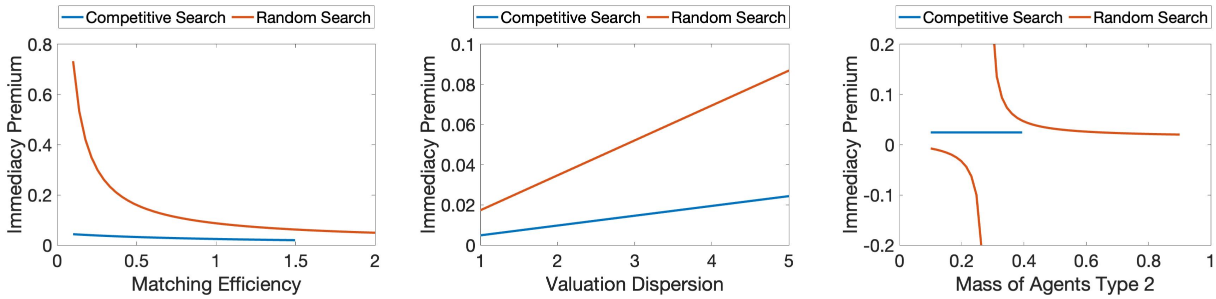


Figure 15: Immediacy Premium for Type 1 Seller in Random and Competitive Search. Immediacy premium for sellers of type 1 for different levels of matching efficiency μ (left), valuation dispersion $|\delta_1 - \delta_2| = |\delta_3 - \delta_2|$ (middle), and the mass of type 2 agents $f(\delta_2)$ (right) in the random (red line) and competitive (blue line) search models.

in Figure 1, and the same intuition applies.