ANNUAL REVIEWS

Annual Review of Economics

The Search Theory of Over-the-Counter Markets

Pierre-Olivier Weill[1,2,3]

[1]Department of Economics, University of California, Los Angeles, California 90095, USA; email: poweill@econ.ucla.edu

[2]National Bureau of Economic Research, Cambridge, Massachusetts 02138, USA

[3]Centre for Economic Policy Research, London EC1V 0DX, United Kingdom

Annu. Rev. Econ. 2020. 12:747–73

First published as a Review in Advance on May 29, 2020

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-091819-124829

JEL codes: G11, G12, G21

Keywords

search frictions, over-the-counter markets, asset pricing

Abstract

I review the recent literature that applies search-and-matching theory to the study of over-the-counter financial markets. I formulate and solve a simple model to illustrate the typical assumptions and economic forces at play in existing work. I then offer thematic tours of the literature and, in the process, discuss avenues for future research.

1. INTRODUCTION

Centralized financial markets are typically organized as limit-order books: all-to-all trading platforms with executable quotes. In contrast, over-the-counter (OTC) markets are harder to categorize. To use Lucas's (2012, p. 272) famous paraphrase of Tolstoy, "all centralized markets are the same, but each OTC market is unique in its own way." However, OTC markets do have common features. In particular, investors trade in small groups and not in all-to-all auctions. For example, they trade with only one dealer or in a request-for-quote (RFQ) auction with a few dealers. There is no pre-trade transparency: Quotes, if any, are not executable and so in principle are up for negotiation. Post-trade transparency may also be limited in the sense that investors may have to negotiate terms of trades under imperfect and often asymmetric information about overall market conditions.

Examples of assets that trade mostly in OTC markets include fixed-income securities, various types of derivatives, repos, and federal funds loans (ISDA 2018, SIFMA 2018). Even for assets that mostly trade in centralized markets, OTC markets may attract an economically significant fraction of volume. For example, traditionally, large blocks of equity have traded bilaterally in the upstairs market (see Harris 2003, chap. 15). Also, in the last decade, the equity market has become extremely fragmented across multiple trading venues, some centralized and some OTC (Tuttle 2014, Hatheway et al. 2017). **Figure 1** breaks down the outstanding supply of publicly traded domestic securities between centralized and OTC markets in 2018.[1] The figure reveals that the value of all domestic assets traded in OTC markets is very large, over USD 50,000 billion. It exceeds the value of assets traded in centralized markets by more than USD 20,000 billion.

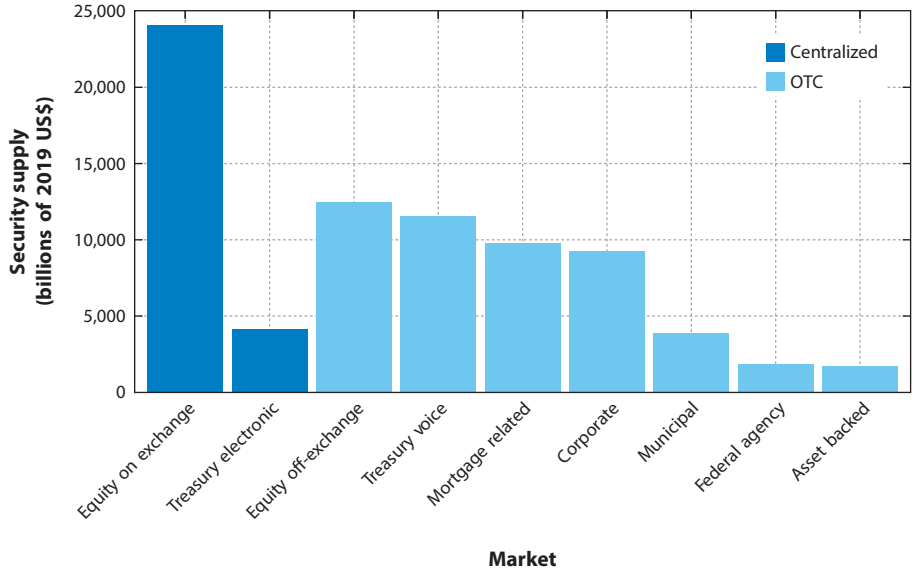

Figure 1

Security supply outstanding in 2018, broken down between centralized and over-the-counter (OTC) markets.

[1] For equities, I use volume share to calculate the fraction of supply traded in centralized and in OTC markets. For treasuries, I use estimates of the volume share of electronic and voice trading for on-the-run and off-the-run securities. See the **Supplemental Appendix** for underlying assumptions and details.

Supplemental Material >

Some have argued that OTC markets are doomed to disappear: Ever improving trading technologies will inevitably make markets more centralized. But the historical record suggests otherwise. Biais & Green (2019) show that, prior to the 1940s, corporate and municipal bonds traded mostly in centralized markets, but the trading volume migrated over time toward OTC markets. Riggs et al. (2018) document that, in the market for credit default swaps, investors can now freely choose between different types of trading mechanisms, some more centralized than others. They find that the most centralized mechanism, a limit-order book, attracts very little volume.

1.1. Why Are We Interested in the Economics of OTC Markets?

Over the last two decades, OTC markets have been at the center of several policy and regulatory debates, all of which underscore the need for rigorous theoretical analyses. For example, in the early 2000s, reporting systems were put in place in the OTC markets for corporate and municipal bonds, with the goal of disseminating price information to the market. This led to an important debate on the economic value of post-trade transparency (see Bessembinder & Maxwell 2008). Several OTC markets were at the core of the 2008 financial crisis, which ignited vigorous policy debates. For example, how should policy makers and regulators respond to the collapse of trading volume in several asset classes traded in OTC markets? Should the government engage in ex-post intervention such as large-scale purchase of mortgage-backed securities (MBS)? Should there be a mandate to implement more centralized and transparent trading mechanisms? Another debate has concerned the appropriate policy response to the buildup of systemic risk in the OTC market for credit derivatives. In particular, should these derivatives be centrally cleared, so as to concentrate counterparty credit risk in central clearing parties?

OTC markets also matter for broader macroeconomic questions. For example, given that most fixed-income securities are traded OTC, these markets impact the cost of capital of firms and sovereigns. For the most part, conventional and unconventional monetary policy is implemented in OTC markets for Treasuries, repos, federal funds loans, and mortgage-related securities. Finally, a commonly held view is that OTC markets have amplified and propagated negative shocks during the financial crisis of 2008.

1.2. Where Does the Theory of OTC Markets Stand?

Most theoretical analyses of OTC markets rely on tools from either search or network theory, and sometimes both at the same time. Two elementary observations motivate these analyses. First, trade is fragmented within small groups of investors: For example, trade may occur bilaterally between two dealers in a voice transaction, or multilaterally between one customer and a small number of dealers in an RFQ auction. Second, an investor is not forced to trade with any group in the OTC marketplace. Instead, an investor can choose with whom and when to trade: For example, a customer can choose between several dealers at any point in time and can choose to trade later with some other dealer. This blurs the distinction between centralized and decentralized markets: The ability to choose with whom and when to trade effectively brings investors together, even in the absence of a centralized all-to-all marketplace.[2]

Search theory provides natural and tractable frameworks to study this choice problem in equilibrium, in particular to derive implications for asset prices, transaction costs, and trading volume,

[2]In fact, one can construct theoretical examples in which a decentralized marketplace achieves the same or nearly the same outcome as a centralized one. Examples include the vanishing friction limits in the dynamic search model of Duffie et al. (2005), the static network–theoretic model of Malamud & Rostek (2017), and the hybrid model of Atkeson et al. (2015).

as well as the supply of intermediation services. Indeed, in a search model, investors trade in small groups, often bilaterally, and they have explicit options to delay trade and find some other counterparty in the market. Since OTC markets are sometimes dominated by a few dealers, network theory is also appealing for a more detailed and precise analysis of strategic interactions among a finite number of investors. One advantage of search over network theory is tractability, as it facilitates the analysis of dynamic trading, of pricing, and of the market response to aggregate shocks. As is often the case, though, it is not wise to be dogmatic about modeling frameworks: Whether a search or a network model is more appropriate ultimately depends on the question at hand and on the economic insights generated. In this review, I rely on my own expertise and focus most of my attention on search-theoretic models.[3]

When applying search theory to financial markets, some scholars experience an uncomfortable stretch of their imagination. They argue that, in practice, investors cannot possibly face an economically meaningful search problem because they either know of their potential counterparties or can quickly learn about them. Several arguments may alleviate this concern. First, even if investors know the full directory of their counterparties, they do not know their trading needs; in particular, finding a counterparty willing to trade large quantities or customized derivative products may take a long time. Second, what makes the search problem economically significant is not the search time per se, but rather the opportunity cost of delaying trade. The magnitude of the price concession that financially distressed institutions are willing to concede (Duffie 2010) suggests that these costs can be quite large. Third, an insight from theory is that the economic significance of search depends not only on the actual time needed to find a counterparty but also on the pricing mechanism. It is well known from the Diamond Paradox (Diamond 1971) that even small time delays can have a very large impact on prices when investors have low bargaining power. Thus, what ultimately matters for valuation is a bargaining-adjusted measure of search times. Fourth, while a search market is clearly an abstraction from the true underlying trade and price-formation mechanism, so is a Walrasian market. The key advantage of the search market is to explain the equilibrium determination of economic phenomena ignored in Walrasian models, such as trading delays, transaction costs, price dispersion, or mismatch. Search models are not inconsistent with Walrasian models; in fact, a desirable property of many search models is to deliver Walrasian prices as frictions vanish.

The remainder of this article is divided in two parts. In Section 2 I analyze a benchmark search model of OTC market. In Section 3, I use this model as a basis to review the existing literature and discuss avenues for future research.

2. A BENCHMARK MODEL

In this section, I formulate and solve a simple benchmark model to illustrate some of the typical assumptions and economic forces discussed in the literature following the pioneering work of Duffie et al. (2005).

---

[3] Models of OTC markets building on the financial network literature have been proposed by Farboodi (2014), Gofman (2014, 2017), Malamud & Rostek (2017), Aymanns et al. (2018), Babus & Farboodi (2018), Babus & Hu (2018), Babus & Kondor (2018), Babus & Parlatore (2018), Eisfeldt et al. (2018), Manea (2018), Wang (2018), and Babus (2019). Some researchers also use a hybrid approach blending elements of search and network models, including Colliard & Demange (2014), Atkeson et al. (2015), Chang & Zhang (2018), Colliard et al. (2018), and Dugast et al. (2019).

2.1. The Setup

The setup builds on Hugonnier et al. (2014), but with a simpler matching and price-setting mechanism.

2.1.1. Assets and agents. Time is continuous and runs forever. There is one asset in positive supply $s \in (0, 1)$. The economy is populated by two types of risk-neutral agents who discount the future at rate $r > 0$: a measure μ_d of dealers, and a measure μ_c of customers normalized to $\mu_c = 1$. Dealers cannot hold any inventory. A customer can hold either zero or one unit of the asset, in which case they derive a utility flow δ. To create a demand for trading assets, I assume that customers' utility flows vary independently over time: Namely, at independent Poisson arrival times with intensity γ, a customer with current utility flow δ draws a new utility flow δ' in a compact interval according to the strictly positive probability density $f(\delta')$.

The assumption that customers derive some time-varying and idiosyncratic utility flow from assets has become standard in the literature, but it may not be obvious to interpret in the context of financial markets. The literature has provided several natural interpretations and microfoundations. First, heterogenous utility flows can reflect agents' heterogenous beliefs about the distribution of asset future payoffs, as proposed by Duffie et al. (2002) and Hugonnier (2012). Second, heterogenous utility flows may arise from agents' heterogenous hedging needs. Imagine for example that agents are risk averse and hold some outside asset that is correlated with the asset under study—think, for example, of a corporate bond dealer who holds a large amount of Treasury bonds. Then large holdings of the outside asset depress the marginal value for the asset under study, generating a small δ, and vice versa [see the empirical study by Newman & Rierson (2003)]. Of course, making this argument precise requires additional work, provided for example by Duffie et al. (2007) and Vayanos & Weill (2008), and especially by the dynamic programming arguments of Praz (2015). Finally, in recent monetary models, customers have heterogenous consumption opportunities, which they can finance by selling assets in OTC markets (see, among others, Geromichalos & Herrenbrueck 2016). The heterogenous consumption opportunities translate into heterogenous indirect utilities for assets.

2.1.2. Market. As first proposed by Duffie et al. (2005), I will assume that the OTC market is semicentralized: Competitive dealers can trade instantly in a centralized interdealer market, but customers must search for dealers in order to trade. Specifically, customers randomly contact dealers at independent Poisson arrival times with intensity λ, and they bargain over the terms of trade in a manner specified below.

Semicentralized markets are realistic: In practice, customers must often trade through intermediaries, and much of the trading volume is made up of either customer-to-dealer trades or dealer-to-dealer trades (see, for example, Atkeson et al. 2013 for credit derivatives). Moreover, interdealer trades may be viewed as more competitive, as they are empirically associated with smaller price dispersion (Li & Schürhoff 2019). The semicentralized market assumption has also proved to be remarkably tractable in extensions of the basic model.

2.2. Bargaining Between Customers and Dealers

Let $V_q(\delta)$ denote the maximum attainable utility of a customer who holds $q \in \{0, 1\}$ units of the asset, with current utility flow δ. Consider a type-δ customer who owns an asset, $q = 1$, and contacts a dealer. Given the $q \in \{0, 1\}$ restriction on asset holdings, the only possible trade is that the customer sells their asset to the dealer. If the trade occurs, the customer receives utility

$B(\delta) + V_0(\delta)$, for some bid price $B(\delta)$ to be determined. If the trade does not occur, the customer continues their search and so receives the utility $V_1(\delta)$. Hence, the net utility of the customer is

$$B(\delta) + V_0(\delta) - V_1(\delta) = B(\delta) - \Delta V(\delta),$$

where $\Delta V(\delta) \equiv V_1(\delta) - V_0(\delta)$ denotes the reservation value of the customer. The net utility of the dealer is, on the other hand,

$$P - B(\delta),$$

where P denotes the interdealer price, to be determined later in equilibrium. Following the literature, I determine the bid price via generalized Nash bargaining,

$$B(\delta) = \arg\max \left(B - \Delta V(\delta)\right)^{1-\theta} (P - B)^{\theta},$$

with respect to B, subject to $B - \Delta V(\delta) \geq 0$, $P - B \geq 0$, and given some $\theta \in [0, 1]$ representing the dealer's bargaining power. One sees that the constraint set is nonempty if and only if $P - \Delta V(\delta) \geq 0$. That is, there are gains from trade between the customer and the dealer if and only if the interdealer price is larger than the customer's reservation value. When the constraint set is nonempty, the first-order necessary and sufficient conditions yield:

$$B(\delta) = \theta \Delta V(\delta) + (1 - \theta)P. \qquad 1.$$

Hence, the bid price lies between the customer's reservation value and the interdealer price. If the dealer has strong bargaining power, $\theta \simeq 1$, then the bid price is close to the customer's reservation value, that is, the price set by a fully discriminating monopsonist. Vice versa, if the customer has strong bargaining power, $\theta \simeq 0$, then the price is competitive.

Proceeding the same way for a customer who does not own an asset, we obtain that there are gains from trade if and only if the customer's reservation value is larger than the interdealer price, $\Delta V(\delta) \geq P$. Moreover, the ask price is given by the same formula as the bid price:

$$A(\delta) = \theta \Delta V(\delta) + (1 - \theta)P. \qquad 2.$$

2.2.1. Price-setting mechanisms. The bid (Equation 1) and the ask (Equation 2) can be viewed as the outcomes of realistic strategic bargaining protocols and price-setting mechanisms in OTC markets.

For example, as is well known, Nash bargaining prices arise when the customer and the dealer engage in a fully specified dynamic bargaining game with alternative offers, as proposed by Rubinstein (1982). One important difference is that the bargaining weight, θ, is now endogenous and depends on other model parameters.

The Nash bargaining payoffs are also obtained if we assume that each customer simultaneously contacts several dealers to receive price quotes, in a game à la Burdett & Judd (1983). This is appealing for at least two reasons. First, since customers can choose between several standing offers, this price-setting mechanism captures in a simple way the possibility of recall. Second, this mechanism also closely resembles an RFQ auction, which is now common in OTC markets: For example, in the corporate bond or credit default swap markets, customers can request quotes from several dealers at the same time (Fermanian et al. 2016, Riggs et al. 2018). To be more specific, let us consider the following stylized representation of an RFQ auction. Assume that each customer sends $n \geq 2$ simultaneous quote requests to randomly chosen dealers, but dealers independently

decline the request with probability $\pi \in (0, 1)$ for exogenous reasons, perhaps because they are busy with other clients [for example, Riggs et al. (2018) show that, in their credit default swaps (CDS) data, dealers do not respond to RFQ about 10% of the time]. Dealers are uncertain about the number of offers received by customers. As a result, the arguments of Burdett & Judd (1983) (see the **Supplemental Appendix**) show that they play a mixed strategy, randomizing their price quotes from a nonatomic distribution. Although the resulting price dispersion is different from what would be obtained with Nash bargaining, the average payoffs are the same. The average transaction price is $\theta \Delta V(\delta) + (1 - \theta)P$, where θ is the customer's probability of receiving just one quote, conditional on receiving some quote. Thus, under this stylized model of RFQ, most predictions of the bilateral trading model remain unchanged (see Lester et al. 2018 for Burdett & Judd pricing under asymmetric information; see Duffie et al. 2017 and Glebkin et al. 2019 for closely related price-setting mechanisms).

Supplemental Material >

The price-setting protocols described above are opaque, in the sense that customers must search for dealers in order to discover or bargain over prices. However, in several OTC markets, regulators have promoted price transparency. A convenient way to model a transparent price-setting protocol is to assume that customers can direct their search toward the price posted by dealers, as in models of competitive search. Lester et al. (2015) provide a detailed analysis of competitive search in the present semicentralized model; Wright et al. (2017) offer a survey of the literature; and Guerrieri & Shimer (2014), Williams (2014), Armenter & Lester (2017), Chang (2018), Chaumont (2018), Li (2018), and Gabrovski & Kospentaris (2020) offer OTC market applications.

2.3. Dynamic Programming

The Hamilton–Jacobi–Bellman (HJB) equations for the customer's maximum attainable utilities are:

$$rV_0(\delta) = \gamma \int \left[V_0(\delta') - V_0(\delta)\right] f(\delta')\, d\delta' + \lambda \max\{\Delta V(\delta) - A(\delta), 0\},$$

$$rV_1(\delta) = \delta + \gamma \int \left[V_1(\delta') - V_1(\delta)\right] f(\delta')\, d\delta' + \lambda \max\{B(\delta) - \Delta V(\delta), 0\}.$$

Taking the difference between the two equations, and using the expression for the bid and the ask prices in Equations 1 and 2 above, we obtain that $\Delta V(\delta) = V_1(\delta) - V_0(\delta)$ solves

$$r\Delta V(\delta) = \delta + \gamma \int \left[\Delta V(\delta') - \Delta V(\delta)\right] f(\delta')\, d\delta' + \lambda(1 - \theta) \max\left\{P - \Delta V(\delta), 0\right\} - \lambda(1 - \theta) \max\left\{\Delta V(\delta) - P, 0\right\}.$$

The left-hand side, $r\Delta V(\delta)$, is the annuitized or flow reservation value. The right-hand side decomposes this flow reservation value into four components. The first component, δ, is the flow utility received by a customer who holds one unit of asset. The second component, $\gamma \int [\Delta V(\delta') - \Delta V(\delta)] f(\delta')\, d\delta$, is the flow of expected net utility of a type change. That is, with intensity γ, the customer draws a new type δ' according to the density $f(\delta')$. This changes the reservation value by $\Delta V(\delta') - \Delta V(\delta)$. The third term, $\lambda(1 - \theta)\max\{P - \Delta V(\delta), 0\}$, is the net utility flow of selling assets. The fourth term, $-\lambda(1 - \theta)\max\{\Delta V(\delta) - P, 0\}$, is minus the net utility flow of purchasing.

A first takeaway from the equation is that reservation values depend on two search options. On the one hand, as shown by the third term, reservation values depend positively on the option

to search for a higher selling price, which increases a customer's willingness to pay. On the other hand, as shown by the fourth term, it depends negatively on the option to search for a lower purchasing price, which naturally decreases a customer's willingness to pay. As will become clear, the price impact of search frictions ultimately depends on the relative magnitude of these two option values.

A second takeaway from the HJB equation is that trading through dealers with intensity λ and bargaining power θ is payoff equivalent to trading directly in the centralized interdealer market, but with a bargaining-adjusted intensity $\lambda(1-\theta)$. This important observation means that what ultimately matters for valuation is not the true physical search intensity, λ, but the bargaining-adjusted search intensity, $\lambda(1-\theta)$. Hence, even when customers contact dealers relatively quickly, search frictions may have a large impact on valuation if dealers have a sufficiently large bargaining power.

2.3.1. A first explicit solution. Now using that $\max\{P-\Delta V(\delta),0\}-\max\{\Delta V(\delta)-P,0\}=P-\Delta V(\delta)$, the HJB equation can be written as

$$r\Delta V(\delta)=\delta+\gamma\int\left[\Delta V(\delta')-\Delta V(\delta)\right]f(\delta')\,d\delta'+\lambda(1-\theta)\left[P-\Delta V(\delta)\right]. \tag{3.}$$

By inspection, one sees that this HJB equation is solved by

$$\Delta V(\delta)=\mathbb{E}\left[\int_0^\tau e^{-rt}\delta_t\,dt+e^{-r\tau}P\,\middle|\,\delta_0=\delta\right], \tag{4.}$$

where τ is exponentially distributed with bargaining-adjusted intensity $\lambda(1-\theta)$, and δ_t is the stochastic utility flow at time $t>0$, which in general differs from $\delta_0=\delta$ because the customer may have switched type. The formula shows that the customer's reservation value is equal to the expected present value of enjoying utility flows until the next contact time with the interdealer market, at which point the value of holding the asset versus the value of not holding it is simply equal to the interdealer price. After a few lines of algebra, one can rewrite this equation as

$$\Delta V(\delta)=\left(1-\mathbb{E}\left[e^{-r\tau}\right]\right)\frac{D(\delta)}{r}+\mathbb{E}\left[e^{-r\tau}\right]P, \tag{5.}$$

where

$$D(\delta)\equiv\frac{\mathbb{E}\left[\int_0^\tau e^{-rt}\delta_t\,dt\,\middle|\,\delta_0=\delta\right]}{\mathbb{E}\left[\int_0^\tau e^{-rt}\,dt\right]}.$$

Equation 5 shows that the reservation value is a convex combination of two terms. To interpret the first term, notice that $D(\delta)$ is the average discounted flow utility of the customer, from $t=0$ to $t=\tau$. The second term is the price. One sees in particular that, if $\lambda(1-\theta)\to\infty$, then $\mathbb{E}[e^{-r\tau}]=\frac{\lambda(1-\theta)}{r+\lambda(1-\theta)}\to 1$, and the reservation value converges to the price. This makes sense: If the customer can contact the interdealer market instantly, then their reservation value is simply the value of selling the asset on that market at price P, regardless of δ. Hence, the only reason the reservation value is different from the price is that there are search and bargaining frictions: A customer who owns an asset must hold it for some time before selling it to the market, with an expected discounted valuation given by the first term in the equation.

2.3.2. A second explicit solution. Let us first solve for the average reservation value by taking expectations on both sides of Equation 3 with respect to the probability density $f(\delta)$. We obtain

$$r\int \Delta V(\delta')f(\delta')\,d\delta' = \int \delta' f(\delta')\,d\delta' + \lambda(1-\theta)P - \lambda(1-\theta)\int \Delta V(\delta')f(\delta')\,d\delta',$$

which implies that

$$\int \Delta V(\delta')f(\delta')\,d\delta' = \frac{\int \delta' f(\delta')\,d\delta' + \lambda(1-\theta)P}{r+\lambda(1-\theta)}.$$

Substituting back into Equation 3 gives, after a few lines of algebra,

$$\Delta V(\delta) = \frac{r}{r+\lambda(1-\theta)}\left[\frac{r+\lambda(1-\theta)}{r+\gamma+\lambda(1-\theta)}\frac{\delta}{r} + \frac{\gamma}{r+\gamma+\lambda(1-\theta)}\int \frac{\delta'}{r} f(\delta')\,d\delta'\right] + \frac{\lambda(1-\theta)}{r+\lambda(1-\theta)}P. \quad 6.$$

Comparing with Equation 5, one obtains that

$$D(\delta) = \frac{r+\lambda(1-\theta)}{r+\gamma+\lambda(1-\theta)}\delta + \frac{\gamma}{r+\gamma+\lambda(1-\theta)}\int \delta' f(\delta')\,d\delta'. \quad 7.$$

This is an explicit expression for the average discounted flow utility of the customer from $t = 0$ to $t = \tau$. It shows, for example, that $D(\delta)$ converges to δ if $r \to \infty$, $\lambda(1-\theta) \to \infty$, or if $\gamma \to 0$. In all these cases, $D(\delta)$ is mostly determined by its initial value, $\delta_0 = \delta$.

2.4. Market Clearing: Allocations

To write the market-clearing condition, I note that, during a small time interval of length h, the measure of customers in contact with dealers is equal to λh. Recall that these customers are sampled independently from the entire population. This has two key implications.

First, within the small group of customers in contact with dealers, the fraction of asset holders is s, the per capita asset supply in the overall customer's population. Therefore, the gross supply of assets during the small time interval of length h is λhs.

Second, within the small group of customers in contact with dealers, the distribution of utility flows is $f(\delta)$, the same as in the customers' overall population. From the analysis of reservation values, it then follows that when a customer meets a dealer, their post-trade holding is $q = 1$ if and only if $\Delta V(\delta) > P$. But since the reservation value is strictly increasing in δ, and is unbounded above and below, this implies that there is some cutoff $\delta^\star$ (possibly outside the support) such that $\Delta V(\delta) > P$ if and only if $\delta > \delta^\star$. The customer with utility type $\delta^\star$ is indifferent between holding the asset or not—that is, $\Delta V(\delta^\star) = P$—and is commonly referred to as the marginal customer. Taken together, the gross demand of assets is $\lambda h \int_{\delta \geq \delta^\star} f(\delta')\,d\delta'$.

Equating gross supply and gross demand, I find that $\delta^\star$ solves

$$s = \int_{\delta \geq \delta^\star} f(\delta')\,d\delta'. \quad 8.$$

This condition reveals that the utility type of the marginal customer, $\delta^\star$, is independent of the search-intensity parameter, λ. However, this does not mean that search frictions do not matter for allocations; it only means that search frictions do not determine who holds the asset within the small group of customers who are currently in contact with dealers. There are other customers below the marginal type, $\delta < \delta^\star$, who hold the asset because they have switched type and have not

been able to contact a dealer. Likewise, there are customers above the marginal type who do not hold.

To see precisely how search frictions matter for allocations, let $\psi_q(\delta)$ denote the density of customers who hold q units of the asset with utility flow δ. In a steady state, the densities must solve the following three equations:

$$\psi_0(\delta) + \psi_1(\delta) = f(\delta), \tag{9.}$$

$$\int \psi_1(\delta')\, d\delta' = s, \tag{10.}$$

$$\lambda \psi_0(\delta) \mathbb{I}_{\{\delta > \delta^\star\}} + \gamma \left[\int \psi_1(\delta')\, d\delta' \right] f(\delta) = \lambda \psi_1(\delta) \mathbb{I}_{\{\delta \leq \delta^\star\}} + \gamma \psi_1(\delta). \tag{11.}$$

Equation 9 states that the total density of type-δ customers must be equal to $f(\delta)$. Equation 10 is a feasibility condition, stating that all assets must be held. Equation 11 equates inflow (on the left side) and outflow (on the right side) into the set of customers with holding $q = 1$ and utility flow δ. (The counterpart of Equation 11 for $q = 0$ is redundant.) Combining the three equations gives:

$$\frac{\psi_1(\delta)}{f(\delta)} = \begin{cases} \frac{\gamma}{\lambda+\gamma} s & \text{if } \delta < \delta^\star \\ \frac{\gamma}{\lambda+\gamma} s + \frac{\lambda}{\lambda+\gamma} & \text{if } \delta \geq \delta^\star \end{cases}.$$

The formula confirms that assets are held by customers of all utility types. However, thanks to reallocation through the search market, customers with utility flow $\delta > \delta^\star$ are more likely to hold assets. In fact, when $\lambda \to \infty$ and the market becomes frictionless, assets are only held by customers with utility flow $\delta > \delta^\star$. When $\lambda \to 0$, assets are held randomly across the utility flow spectrum. For intermediate values of λ, the distribution is a convex combination of the $\lambda \to \infty$ frictionless allocation and of the $\lambda = 0$ random allocation with a convex weight $\lambda/(\lambda + \gamma)$. The magnitude of the weight is fully determined by the ratio λ/γ. For instance, if $\gamma \simeq 0$, the allocation converges to the frictionless allocation, even though $\lambda < \infty$. This is because, in this case, customers never change utility flow. As a result, after their first contact time with dealers, a customer with $\delta > \delta^\star$ will hold an asset forever.

2.5. Market Clearing: Prices

I turn to an analysis of prices. From the previous section, it follows that, conditional on contacting dealers, a customer finds it optimal to hold the asset if and only if $\delta > \delta^\star$. Therefore, we have that $P = \Delta V(\delta^\star)$. Plugging this equality into Equation 5 evaluated at $\delta = \delta^\star$, we obtain

$$P = \frac{D(\delta^\star)}{r} = \frac{1}{r} \left(\frac{r + \lambda(1-\theta)}{r + \gamma + \lambda(1-\theta)} \delta^\star + \frac{\gamma}{r + \gamma + \lambda(1-\theta)} \int \delta' f(\delta')\, d\delta' \right), \tag{12.}$$

where the second equality follows from Equation 7.

To characterize the impact of search frictions on the asset price, it is useful to analyze two limit cases, when frictions are very large, $\lambda(1-\theta) \to 0$, and when frictions are very small, $\lambda(1-\theta) \to \infty$:

$$\lim_{\lambda(1-\theta)\to 0} P = \mathbb{E}\left[\int_0^\infty e^{-rt} \delta_t\, dt \,\middle|\, \delta_0 = \delta^\star \right] = \frac{1}{r} \left(\frac{r}{r+\gamma} \delta^\star + \frac{\gamma}{r+\gamma} \int \delta'\, f(\delta')\, d\delta' \right), \tag{13.}$$

$$\lim_{\lambda(1-\theta)\to \infty} P = \frac{\delta^\star}{r}. \tag{14.}$$

Equation 13 is the price when customers expect that it will take them an infinitely long bargaining-adjusted time to contact dealers. The expression reveals that the price in this case is equal to the autarky, or buy-and-hold, utility of the marginal customer. One can easily verify that it also coincides with the asset price if a competitive market opens at time $t = 0$ only and never after.

Equation 14 is the price when customers can trade instantly. In that case, the price is equal to the present value of the utility flow of a hypothetical customer who remains the marginal type forever. This is so even though, according to our specification of the utility flow process, no customer actually remains marginal forever. Indeed, whenever a customer's utility type jumps below $\delta^\star$, they find it optimal to sell immediately, and vice versa when a customer's utility flow jumps above $\delta^\star$. As a result, the market price capitalizes the flow utility of a hypothetical customer who is always indifferent between buying and selling.

Comparing Equations 13 and 14, one sees that the $\lambda(1-\theta) \to 0$ price is smaller than $\lambda(1-\theta) \to \infty$ if $\delta^\star > \int \delta' f(\delta')\, d\delta'$, and greater otherwise. Correspondingly, when $\lambda(1-\theta)$ increases from zero to infinity, the price increases if $\delta^\star > \int \delta' f(\delta')\, d\delta'$, and it decreases otherwise.

To understand this finding, recall that reservation values depend positively on the search option to sell and negatively on the option to buy. When $\lambda(1-\theta)$ increases, both options become more valuable, affecting reservation values in opposite ways. But when $\delta^\star > \int \delta' f(\delta')\, d\delta'$, the marginal customer expects their utility flow to fall over time, so the option to sell if they own has a larger value than the option to buy if they do not. Hence, as $\lambda(1-\theta)$ increases, the positive effect of the option to sell dominates the negative effect of the option to buy, and the price increases.

Feldhütter (2012) argues that this observation helps to empirically identify times of strong selling pressures in the market. To understand his argument within the benchmark model, note that the condition $\delta^\star < \int \delta' f(\delta')\, d\delta'$ indicates a strong selling pressure: Indeed, Equation 8 reveals that the condition holds either when supply is large or when demand is low. As argued above, under this strong selling pressure condition, the option value to sell is less valuable than the option to buy, and so an increase in $\lambda(1-\theta)$ decreases prices. Hence, in a cross-section of investors, one should expect that the most sophisticated ones—that is, those with a high $\lambda(1-\theta)$—trade at lower prices than the less sophisticated ones. The opposite is true when the selling pressure subsides. Hence, the sign of the price differential between sophisticated and less sophisticated investors helps identify times of strong selling pressure in the market.

2.5.1. Decomposing the yield spread. Imagine that the asset is a fixed-income security, such as a municipal or corporate bond. Then, the natural empirical counterpart of the discount is the liquidity yield spread. That is, let us define the yield on the asset to be $r + \ell$, where ℓ is the spread over the risk-free rate that makes the present value of the marginal customer's utility flow equal to the price,

$$P = \frac{\delta^\star}{r+\ell}.$$

Notice the desirable property that $\ell = 0$ when there is no friction. Then, using Equation 12, we obtain that

$$\frac{\ell}{r+\ell} = \frac{\gamma}{r+\gamma+\lambda(1-\theta)}\left(1 - \frac{\int \delta' f(\delta')\, d\delta'}{\delta^\star}\right).$$

This equation confirms that the liquidity yield spread is positive if and only if $\delta^\star > \int \delta' f(\delta')\, d\delta'$, that is, if and only if a type change reduces the marginal customer's utility flow for the asset on average, which may be interpreted as financial distress. The equation also shows that the yield

spread depends jointly on several parameters: It is increasing in the frequency of financial distress, as measured by γ; in the average distress cost, as measured by $1 - \int \delta' f(\delta')\, d\delta' / \delta^\star$; in the size of search frictions, as measured by $1/\lambda$; and in the market power of dealers, as measured by θ. Gavazza (2016) and Hugonnier et al. (2020) discuss how bargaining power and distress cost are separately identified by the level of prices (yield spread) and transaction costs (bid-ask spread).

3. THEMATIC TOURS OF THE LITERATURE

This section provides thematic tours of the literature. Each tour focuses on a collection of papers sharing a methodological or substantive theme. The tours often cross paths: Papers may appear multiple times if they fit in different themes.

3.1. Entry

A number of authors have considered entry decisions in OTC markets. Consider for example the entry decision of dealers. Assume that, if the total measure of dealers in the market is μ_d, then customers' search intensity is given by an increasing and concave function $\lambda(\mu_d)$ satisfying Inada conditions. Hence, when the measure of dealers increase, the semicentralized market creates more contacts with customers but becomes more congested. Flow profit per dealer can be shown to be

$$\frac{\lambda(\mu_d)\gamma s(1-s)}{\mu_d\,(\gamma + \lambda(\mu_d))} \frac{\theta\left(\mathbb{E}\left[\delta' - \delta^\star \mid \delta' > \delta^\star\right] + \mathbb{E}\left[\delta^\star - \delta' \mid \delta' < \delta^\star\right]\right)}{r + \gamma + \lambda(\mu_d)(1-\theta)},$$

where all expectations are taken with respect to the distribution $f(\delta')$. The first term is the volume per dealer and the second term is the average bid-ask spread. One sees that the profit per dealer is decreasing in μ_d, for two reasons. First, although the total volume increases with μ_d, the volume per dealer goes down because $\lambda(\mu_d)$ is concave. Second, the entry of dealers improves the outside option of customers and reduces the average bid-ask spread. Clearly, there is a unique value of μ_d for which flow profits and entry costs are equalized. One can show that, relative to a utilitarian planner's benchmark, dealers enter too much if θ is too large and too little if θ is too small, as in many other search-and-matching models.

In his structural study, Gavazza (2016) finds excessive entry of dealers in the market for commercial aircraft. Vayanos & Wang (2007), Afonso (2010), Rocheteau & Weill (2011), and Sambalaibat (2015) consider the entry of customers in OTC markets. Atkeson et al. (2015) consider both entry and exit of investors with heterogenous valuation, and they show that, depending on their preferences, some investors endogenously choose to enter and assume the role of dealers, while others choose to enter and assume the role of customers. They show that, under natural conditions, too many investors enter to assume the role of dealers, and too little to assume the role of customers. In work by Farboodi et al. (2018b), ex-ante identical investors endogenously choose different search intensities upon entry, resulting in an equilibrium market structure similar to the one exogenously assumed in the benchmark model. Finally, models of competitive search typically impose a free entry condition as well, but they typically obtain efficiency under symmetric information.

3.2. Unrestricted Asset Holdings

In the benchmark model, we assumed for simplicity that investors can hold either one or zero units of this asset. Gârleanu (2009) and Lagos & Rocheteau (2009) have shown independently that the model can be generalized to the case of unrestricted asset holdings. In particular, Lagos

& Rocheteau (2009) consider the following specification of preferences: They assume that an investor with utility type δ who holds q units of the asset derives a flow utility equal to $\delta u(q)$, and the utility function satisfies Inada conditions.[4] More recent models with unrestricted holdings but fully decentralized search markets have been proposed by Cujean & Praz (2013), Afonso & Lagos (2015b), Üslü (2019), and Üslü & Velioğlu (2019).

With unrestricted asset holdings, the pricing formula of the benchmark model remains valid at the margin, with a utility flow equal to $\delta u'(q)$ instead of just δ. The difference, of course, is that the asset holding q is now endogenous and depends on the magnitude of the frictions. For example, customers endogenously respond to a reduction in frictions by increasing the size of their trades—in this sense, they demand more transaction services from dealers. This can have an important impact on the model's prediction. Consider for example the question of dealers' entry in Lagos & Rocheteau's (2007) article. In the benchmark model, we have seen that per-dealer profits are decreasing in the measure of dealers, so that the entry equilibrium is uniquely determined. In the model with unrestricted asset holdings, there is an additional effect: With more dealers, customers meet dealers more quickly and so are willing to trade larger quantities. This effect can make per-dealer profit increase in the measure of dealers, creating strategic complementarities in entry decisions and multiple equilibria.

3.3. Alternative Assumptions About the Matching Technology

The matching function is defined as the mapping between the measures of various market participants and the total flow of meetings between them. For example, in the semicentralized market, if the measure of customers is $\mu_c = 1$ and the measure of dealers is μ_d, then the flow of meetings is simply $\lambda\mu_c$. More generally, the matching function could depend positively on the measure of both customers and dealers in arbitrary ways. It is customary in the search literature to assume that the matching function has constant returns to scale; that is, scaling up all measures by the same constant scales the flow of meetings linearly, but it does not impact search times. Put differently, liquidity does not depend on market size. In labor market applications, this assumption appears to be supported by empirical evidence (Petrongolo & Pissarides 2001). In finance, however, it is not unreasonable to believe that larger markets are more liquid—for example, empirically, there typically exists a positive relationship between issue size and liquidity measures. Correspondingly, some of the literature has considered increasing returns in matching (Pagano 1989, Vayanos & Wang 2007, Vayanos & Weill 2008, Weill 2008, Sambalaibat 2015, Shen et al. 2015).

An & Zheng (2018) and An (2019) use a different model of the matching process, in the tradition of the stock flow–matching literature (Coles & Smith 1992). According to this model, trading delays do not arise from search frictions, but rather because customers only demand specific assets. Assume, for example, that there is a flow γ_b of customers who turn into buyers and instantly contact the market. Each buyer is only willing to purchase a finite collection of assets, drawn at random in the continuum according to a Poisson distribution with parameter ρ, independently across buyers. As a result, the demand for any particular asset will take time to materialize—for example, if there is just one single asset for sale, then a customer willing to buy this asset will arrive at

[4]Without the Inada conditions, the benchmark model with $q \in \{0, 1\}$ asset holding is a special case of the model with unrestricted holdings: Indeed, equilibria in which holdings are restricted to $q \in \{0, 1\}$ remain equilibria with unrestricted holdings, as long as utility is Leontief $u(q) = \min\{q, 1\}$ and $q \geq 0$. This is because, in this case, investors find it optimal to hold either $q = 0$ or $q = 1$. With Inada conditions, the benchmark model is no longer a special case but it can be approximated with arbitrary accuracy by appropriate choices of $u(q)$ (see Biais et al. 2014).

an exponentially distributed time with parameter $\rho\gamma_b$. But this means that, in equilibrium, a stock of buyers and sellers is waiting in the market for suitable counterparties to arrive. In this context, An & Zheng (2018) and An (2019) show that imperfectly competitive dealers have incentives to hold inventories in spite of high holding costs. In their equilibrium, dealers source assets in two ways: from their inventory or from the set of sellers waiting for a suitable counterparty. This is appealing because it resembles risky-principal and riskless-principal trade in practice.

3.4. Fully Decentralized OTC Markets

The semicentralized market assumption is based on the view that interdealer frictions are smaller than customer-to-dealer frictions. However, much of the micro data about OTC market concerns interdealer trades. For example, in the corporate bond data from the Trade Reporting and Compliance Engine (TRACE), anonymized identifiers track the trades of particular dealers, but no such identifiers are provided for customers. Hence, for better or worse, the interdealer market is our main empirical laboratory for studying OTC market frictions. This requires going beyond semicentralized markets and formulating models that are fully decentralized, in the sense that all trades, including interdealers, occur in a decentralized market. Moreover, the empirical evidence about interdealer trades suggests that there is substantial heterogeneity between dealers in terms of their trading volume, their markups, whether they tend to trade more often with customers or with other dealers, and so on.

However, solving models with heterogenous dealers in fully decentralized markets turns out to be quite complex. To understand why, assume as in the benchmark model that agents are heterogenous in terms of their utility flow, but that they must search for counterparties in a fully decentralized market. Then, the reservation value $\Delta V(\delta)$ depends on their expected terms of trade, which in turn depend on the distribution of reservation values across potential counterparties, determined by $\psi_1(\delta)$ and $\psi_0(\delta)$. Hence, one must wrestle with a potentially high-dimensional fixed-point problem: characterizing the two-way feedback between this distribution and agents' optimal trading decisions.[5] An active recent literature has developed methods to solve this problem, including work by Afonso & Lagos (2011), Colliard & Demange (2014), Hugonnier et al. (2014, 2020), Atkeson et al. (2015), Shen et al. (2015), Bethune et al. (2018), Chang & Zhang (2018), Farboodi et al. (2018a,b), Tse & Xu (2018), Neklyudov (2019), Üslü (2019), Üslü & Velioğlu (2019), and Yang & Zeng (2019). Because of their rich implications and because they apply to interdealer trades, these models take the literature one step closer to structural estimation and ex-ante policy evaluation. In particular, Liu (2020) has recently offered a structural estimation of the model by Hugonnier et al. (2020) with endogenous search intensity. But more progress is needed in this important area of research.

A key theoretical insight in this literature is that, with heterogenous agents and fully decentralized markets, some agents endogenously emerge as dealers and others as customers. To understand why, consider our benchmark model in which agents have heterogenous utility flows δ (a similar argument applies to other types of heterogeneity). In a fully decentralized market, agents have heterogenous reservation values and draw counterparties at random from the entire population. Agents whose reservation value $\Delta V(\delta)$ is closer to the economy wide median are equally likely to meet with agents who have either higher or lower reservation value, and so they are equally likely to be buyers or sellers. In equilibrium, these median agents buy and sell repeatedly, so that their

[5] Semicentralized markets are simpler because the distribution of reservation values across counterparties is degenerate: All customers' counterparties are dealers, and all dealers trade in a centralized market, equalizing their reservation values to the price P.

gross trade exceeds their net trade. In this sense, they endogenously emerge as intermediaries or dealers. Agents with extreme reservation values are more likely to trade in just one direction, buy or sell, and so they endogenously emerge as customers.[6]

While different types of heterogeneity can generate similar patterns of endogenous intermediation, they have different economic implications. Perhaps the simplest example would be intermediation arising because of differences in trading speed (Farboodi et al. 2018b) versus differences in rent-extraction ability (Farboodi et al. 2018a). The former case leads to strictly efficient intermediation, while the latter does not. Dugast et al. (2019) show that heterogeneity in private values generates OTC markets that are too small relative to centralized markets, while heterogeneity in trading technology has the opposite implication.

One important open question for this theoretical literature is empirical and quantitative: how to determine the relative importance of different types of heterogeneity. Dugast et al. (2019) and Üslü (2019) observe, for example, that the patterns of net and gross trade volume in the cross-section can help distinguish between heterogeneity in flow utility and heterogeneity in trading technology. However, more work remains to be done by studying the qualitative and quantitative implications of different types of heterogeneity for a broader set of market outcomes.

3.5. Dynamic Market Response to Shocks

A substantial body of empirical evidence suggests that two important dimensions of illiquidity are the temporary price impact of supply shocks and the extent to which intermediaries step in to mitigate these supply shocks by taking inventories. The temporary price impact of supply shocks has been studied by Duffie et al. (2007), Feldhütter (2012), Trejos & Wright (2016), and Akın & Platt (2019). The manner in which dealers endogenously respond to supply shocks by accumulating inventories and providing liquidity to customers is analyzed by Weill (2004, 2007, 2011), Lagos et al. (2011), and Di Maggio (2013). Technically, studying this question requires relaxing the constraint that dealers cannot hold inventories. While this constraint does not bind in the steady state of the benchmark model, it turns out to matter out of the steady state. Dealers have incentives to accumulate inventories when the supply shock hits, because they anticipate that they will be able to resell them more quickly than customers when the shock subsides. Models in this vein shed light on the recent policy debate regarding the potentially negative impact of post-crisis regulation on bond market liquidity (Bao et al. 2018, Bessembinder et al. 2018, Dick-Nielsen & Rossi 2019). Empirical work has highlighted that the answer depends on the liquidity measure considered. For example, some authors have argued that the bid-ask spread has not increased. However, in response to supply shocks such as delisting or downgrading, dealers accumulate less inventories than before, and they earn higher returns from liquidity provision. This prediction is entirely consistent with search-based models of dealer's liquidity provision—for example, a simple extension of Weill's (2007) work in which dealers have positive bargaining power. According to the model, the reason the bid-ask spread does not go up is that inventory costs move the bid and the ask in the same direction: Dealers are less willing to buy, which pushes the bid down, but by the same token they are more willing to sell, which pushes the ask down. In this model,

[6]The argument in this paragraph relies on matching being random. However, Chang & Zhang (2018) show that this is not necessary. They consider a dynamic bilateral market in which agents are ex ante identical and have unobservable valuation. In contrast with the above cited literature, meetings are not random: Agents can choose with whom to match. They show that endogenous intermediation arises here, too, as an efficient coordination mechanism to dynamically reduce misallocation. Gabrosvski & Kospentaris (2020) also show that intermediation can arise in the absence of random matching in the context of a competitive search model.

inventory costs do not matter for bid-ask spreads, but they do matter for dealers' returns and for the amount of liquidity they supply.

Another context in which this question has been of interest is that of market freeze. This issue became particularly salient during the Great Financial Crisis of 2008, when trading volume collapsed precisely in markets that were plagued by a sudden increase in asymmetric information: namely, in markets for asset-backed securities in which investors had serious concerns about collateral quality. Camargo & Lester (2014) and Chiu & Koeppl (2016) study theoretically how the market recovers from such asymmetric information shocks and have characterized welfare-improving policy interventions, while Zou (2019) considers dynamic information acquisition decisions.

3.6. Asymmetric Information

In the benchmark model I assumed that prices are set under symmetric information about flow valuation and asset fundamental value. This is clearly a strong assumption. In fact, OTC markets are typically considered opaque, with pervasive asymmetric information problems: about customers and dealers' private values, about outside trading opportunities, about assets' fundamental value, and about aggregate order flow. Many policy interventions and regulations have aimed to alleviate asymmetric information problems. To understand the effect of opaqueness and the potential policy responses, many authors have introduced asymmetric information into search models of OTC markets.

Consider first the case of asymmetric information about private flow valuation. Zhu (2011) considers a sequential search problem of a customer contacting a finite number of dealers, with asymmetric information about valuation but also about the history of past quotes from other dealers. In this context, Zhu shows that repeat contact with the same dealer signals that the customer did not receive good quotes from others and therefore prompts the dealer to offer worse terms. Zhang (2017) considers long-term relationships when dealers do not observe customers' flow valuations. He shows, among other results, that dealers use delays to screen customers, creating an endogenous distortion above and beyond the delays that may be created by search. Cujean & Praz (2013) consider a trading mechanism that can accommodate private information about private value in a fully decentralized market and study the impact of increased transparency.

The literature has considered asymmetric information about common value as well. One typical assumption, in line with the lemon market literature, is that all assets in the market are finely differentiated: That is, the value of each asset is drawn independently according to some common distribution and is known by the seller but not by the buyer. This is relevant to the Great Financial Crisis of 2008. One may argue that in the mortgage-based security market, securities had finely differentiated collateral pools, and that investors were asymmetrically informed about the quality of these pools. Part of the literature seeks to explain the manner in which such asymmetric information reduces liquidity, leads to market freezes or fire sales, and creates room for welfare-improving policy interventions. Research in this area includes work by Guerrieri & Shimer (2014, 2018) and Chang (2018), who use and extend the competitive search framework of Guerrieri et al. (2010) to asset markets (see also Williams 2014, Li 2018). Camargo & Lester (2014) and Chiu & Koeppl (2016) consider dynamics following an adverse selection shock. Maurin (2018) shows that such a setting may display endogenous fluctuation in liquidity. Zou (2019) studies nonstationary equilibria with information acquisition decisions. Lester et al. (2019) extend the Burdett & Judd (1983) pricing mechanism to the case of asymmetric information. They show that, in order to evaluate the impact of policy or regulatory proposals on OTC markets, it is crucial to simultaneously account for imperfect competition and asymmetric information. For example, increasing competition in an OTC market may be either welfare improving or reducing, depending on the degree of asymmetric information.

To address the question of informational efficiency, the market microstructure literature classically considers asymmetric information in a market for a single asset, not many finely differentiated ones. But this creates an important technical difficulty for search models. Indeed, in a fully decentralized market, investors learn about the asset value via their idiosyncratic matching and trading histories. Hence, to characterize equilibrium outcomes, one needs ways to mathematically track investors' increasingly heterogenous learning histories and to characterize the dynamic of the associated distribution of posterior beliefs. Wolinsky (1990) and recently Lauermann et al. (2018) have solved this problem by assuming that new traders continuously enter the economy, making the distribution of posterior beliefs stationary. Blouin & Serrano (2001) consider nonstationary dynamics in the setup of Wolinsky. Duffie & Manso (2007) and Duffie et al. (2009) use convolution methods, assuming that agents fully reveal their information in bilateral meetings. Amador & Weill (2006) consider noisy revelation in a Gaussian setting. Matters become more complicated when one imposes the restriction that signals are generated by trades. Duffie et al. (2014) consider a double auction setting that endogenously results in full revelation. Golosov et al. (2014) characterize long-run outcomes in a setting where trade reveals only partial information.

Lester et al. (2019) consider a semicentralized market in which customers are perfectly informed but dealers are not. One key insight of their analysis is that a decrease in search frictions can increase asymmetric information and, correspondingly, the bid-ask spread. To gain some intuition, consider the following heuristic argument in the benchmark model. Suppose that customers' flow valuation is of the form $v + \varepsilon_i$, where v is a common value component and ε_i is an idiosyncratic component. Consider a match between a customer who is informed about v and ε_i and a dealer who is not. Suppose that, by trading, the dealer can learn the reservation value, $\Delta V(\delta)$. Since the dealer already knows the price, P, this signal is observationally equivalent to $D(\delta)$. Substituting $\delta = v + \varepsilon_i$ into Equation 7, we then obtain that the signal acquired by the dealer is of the form $v + \frac{r+\lambda(1-\theta)}{r+\gamma+\lambda(1-\theta)}\varepsilon_i$, reflecting the average valuation of a customer between two contact times with dealers. The common component has a weight of one, because it never changes between two contact times. The idiosyncratic component may change, so it has a smaller weight. But when $\lambda(1-\theta)$ increases, and correspondingly search frictions decrease, the idiosyncratic component is less likely to change between two contact times, and so it has higher weight. This increases the noise in the signal, reduces the amount of information about v revealed by the trade, leads to more asymmetric information, and widens the spread.

Other approaches to common value–asymmetric information include those by Duffie et al. (2017), who assume that dealers are uncertain about dealers' common market-making cost, and by Brancaccio et al. (2017), who consider asymmetric information about the aggregate customer's order flow.

3.7. Multiple Assets

Much of the empirical evidence about the impact of liquidity on asset prices is cross-sectional—for example, the evidence that risk-adjusted returns are empirically related to liquidity proxies such as turnover, volume, or bid-ask spread (see, for example Amihud et al. 2005, chap. 3), or the evidence that liquidity is related to violation of no-arbitrage relationships (Amihud & Mendelson 1991). Comparative statics of single-asset models are not appropriate to interpret such evidence; instead, one needs to formulate models in which multiple assets are traded. In particular, one needs to explain why an investor's ability to choose between payoff-equivalent assets does not undo an arbitrage relationship or equalize risk-adjusted return differentials.

In work by Vayanos & Wang (2007) and Weill (2008), investors must choose between markets for indivisible assets. Vayanos & Wang assume that assets are homogenous but investors differ in

their investment horizons, while Weill assumes that investors are homogenous but assets differ in supply and possibly other characteristics. In equilibrium, investors are indifferent between asset markets. With increasing returns in matching, large supply assets have higher turnover, lower search times, and higher prices. Therefore, in the cross-section, there is an increasing relationship between supply and price, the opposite of what would be obtained via comparative statics in a single-asset model. Milbradt (2017) also considers asset heterogeneity, but in a semicentralized market with endogenous customers' search intensity. A key innovation is that asset types are changing, possibly stochastically, over time. A natural example would be the time to maturity, the credit rating, or the distance to default of a bond. Milbradt offers new tools and predictions for the joint relationship between volume, prices, and volatility in the cross-section of asset characteristics.

Vayanos & Weill (2008) study the on-the-run phenomenon—that is, the price differential between recent and old Treasury issues with identical time to maturity. They consider a model with two payoff-identical assets and two search markets: one spot market for buying and selling assets and one repo market for borrowing and lending the assets. Investors are not constrained to choose between markets: They receive trading opportunities for either asset. In equilibrium, short-sellers endogenously coordinate to borrow the same asset. Indeed, short-selling activity increases the turnover, reduces search time, and makes it easier to locate the asset when unwinding a short position. The model decomposes the yield spread between its various components and explains what makes arbitrage unprofitable. Sambalaibat (2015) shows that the introduction of CDS can create liquidity spillover in the underlying bond market. This is because the entry of buyers of protection effectively increases supply. With increasing returns in matching, this increases entry on the other side of the market more than one-to-one, so that bonds become easier to sell and have a higher price. These predictions are supported by evidence from the ban on naked CDS trading during the European debt crisis.

In the papers reviewed above, agents are assumed to be risk neutral and are restricted to hold either zero or one unit of the many assets traded in equilibrium. While analytically convenient, these assumptions are at odds with classical portfolio theory, in which diversification benefits generate a demand for broad portfolios. This poses a clear challenge to search-based models, in which it is typically difficult to study how asset demand is shaped by both diversification and liquidity concerns. In a recent paper, Üslü & Velioğlu (2019) integrate classical portfolio choice and asset-pricing theory within a search model of OTC markets. They assume that agents derive a mean-variance utility flow from asset portfolios and receive random opportunities to trade specific assets. Üslü & Velioğlu take as given asset-specific search times and risk exposures, and they derive analytically cross-sectional prices, volume, and price impact. They test the model's key predictions in the corporate bond market.

3.8. OTC Versus Centralized Trade

Why do some investors and assets trade in OTC instead of centralized markets? Answering this question matters a great deal for the ex-ante evaluation of policies that promote centralized market trade. Indeed, the unintended consequences of these policies are likely to depend on the underlying drivers of demand for OTC versus centralized markets.

The literature on this important topic is growing, and it often moves away from the search paradigm. This gives scholars more flexibility to introduce a variety of asymmetric information frictions. A branch of the literature provides comparative static analysis, comparing outcomes across exogenously given market structures, some centralized and other decentralized, or across decentralized structures with varying levels of frictions. Examples include work by Biais (1993), Geromichalos & Herrenbrueck (2016), Malamud & Rostek (2017), Colliard et al. (2018),

Liu et al. (2018), Li & Song (2019), Vogel (2019), and Glode & Opp (2020). Another branch of the literature formalizes the demand for OTC versus centralized market by studying the sorting of heterogenous investors across markets. Examples include work by Yavaş (1992), Gehrig (1993), Rust & Hall (2003), Miao (2005), Yoon (2017), Lee & Wang (2018), and Dugast et al. (2019). A few papers have explored the manner in which decentralized market structures are endogenously offered in equilibrium due to information and price-setting frictions and may dominate a centralized exchange (see Kawakami 2017, Babus & Parlatore 2018, Cespa & Vives 2018, and Farboodi et al. 2018b). In these works, authors offer economic explanations for the emergence of decentralized trade, with many insights about the impact of policies. For the most part, however, they take some element of the price-setting mechanism as given. Thus, it is typically unclear whether decentralized trade emerges because of the price-setting mechanism assumption or because of some primitive informational frictions. What the literature is perhaps missing is a systematic mechanism and information design approach to the emergence of decentralized trading structures.

3.9. Multiplicity and Fragility

Search-and-matching models are a natural framework to explore multiple equilibria in which liquidity begets liquidity. As is well known after Diamond (1982), this can typically be achieved via increasing returns in matching (Vayanos & Wang 2007, Sambalaibat 2018, Vayanos & Weill 2008) or other mechanisms (Lagos & Rocheteau 2007, Chiu & Koeppl 2016, Sultanum 2018, Nosal et al. 2019, Yang & Zeng 2019). Multiple equilibria are appealing for at least two reasons. First, they can help explain phenomena that are difficult to relate to some underlying fundamental characteristics—for example, why payoff equivalent assets have different prices, or why observationally similar intermediaries play different roles. Second, they provide natural frameworks to address fragility and to uncover feedback loops leading to the deterioration of market liquidity.

3.10. Repeat Trade and Relationships

A common criticism of the literature following Duffie et al. (2005) is that counterparties do not trade repeatedly. For example, in the benchmark model, customers never contact the same dealer in the μ_d continuum more than once. This is strongly at odds with empirical evidence (see Afonso et al. 2014, among others). One way to address this criticism is to be explicit about imperfect competition among a finite number of dealers: Indeed, even with search and random matching, atomic dealers will be contacted repeatedly. Repeat contacts are not just mechanical: They do affect the terms of trade (Zhu 2011, An 2019). However, repeat contacts are different from relationships, which create economic value. While the study of long-term relationships is relatively undeveloped in OTC market contexts (with some notable exception, such as Zhang 2017, Sambalaibat 2018, Hendershott et al. 2020), they have been explored in the broader search literature. In particular, the canonical labor market model of Mortensen & Pissarides (1994) addresses the formation of relationship between firms and workers. One perhaps important conceptual distinction is that relationships tend to be exclusive in labor markets but are nonexclusive in OTC markets: Customers may establish simultaneous relationships with several dealers (Afonso et al. 2014).

3.11. Search Models of Centralized Exchange

In the last decades, both individual investors and exchanges have made considerable investments to increase trading speed. Because search theory is built on the premise that speed is scarce in the sense that trade is not instantaneous, it offers a natural framework to study the demand and

supply for speed. Pagnotta & Philippon (2018) consider competition between exchanges that allow investors to trade in semicentralized search markets, formally similar to the benchmark model discussed in Section 2. Because customers are heterogenous with respect to the distribution from which they draw preference shocks, they have different marginal valuations for speed. To see this, interpret the benchmark model with $\theta = 0$ as an exchange offering a level of speed λ. In that market, the ex-ante flow welfare of customers is

$$W(\lambda, f) = \int \delta \psi_1(\delta)\, d\delta = s\mathbb{E}\left[\delta'\right] + \frac{\lambda s}{\lambda + \gamma}\left(\mathbb{E}\left[\delta' \mid \delta' \geq \delta^\star\right] - \mathbb{E}\left[\delta'\right]\right),$$

where expectations are with respect to the distribution of utility flows, $f(\delta')$. Under the assumptions that $s = 1/2$ and that $f(\delta')$ is symmetric, one easily sees that $\partial W/\partial\lambda$ increases if the distribution $f(\delta')$ becomes riskier in the sense of second-order stochastic dominance. This is because of the second term in the expression of $W(\lambda, f)$: With a higher speed, customers can acquire assets more quickly upon drawing a flow valuation above $\delta^\star$. Therefore, there is complementarity between the trading speed of an exchange and the riskiness of an investor's flow utility distribution, $f(\delta')$. This implies that Bertrand competition between exchanges generates horizontal differentiation in trading speed: Exchanges offer higher speed to investors with riskier $f(\delta')$. With fixed costs of establishing exchanges, such differentiation is welfare reducing.

Limit-order markets can also be studied using tools from search theory. Rosu (2009) considers the limit-order pricing strategy of patient investors who wait for the Poisson arrival of impatient traders on the other side of the market. Biais & Weill (2009) and Biais et al. (2014) start from the view that limit orders represent the trading interest of "absent traders . . . while they attend to business elsewhere" (Harris 2003, pp. 77–78). They consider a version of the semicentralized model in which the parameter λ now represents the intensity with which a given customer monitors the market. The key difference with the semicentralized model is that customers are now allowed to trade in two ways: immediately with a market order, or potentially with a delay by leaving a limit order. Dugast (2018) studies limit-order market dynamics following the arrival of unscheduled news.

3.12. Broader Implications of OTC Market Frictions

Much of the literature reviewed so far focuses on partial equilibrium analyses, examining one specific OTC market in isolation from the rest of the economy. However, an active branch of the literature offers general equilibrium analyses studying the two-way feedback between OTC markets and other economic decisions (e.g., a firm's decision to default or the central bank implementation of monetary policy), other markets (e.g., the primary market for bond issuance), or the economy as a whole. This is an important area of research: It shows how the precise modeling of microeconomic frictions in OTC markets can matter for the study of broader economic questions.

The new monetarist tradition, which predates work by Duffie et al. (2005), has considered general equilibrium models with decentralized markets in which goods are exchanged for money and other assets. In the interest of space, I will not review this branch of the literature in depth but instead refer the reader to Lagos et al. (2017) for a survey, Nosal & Rocheteau (2011) for a book-length discussion, and Trejos & Wright (2016) for a comparison with the OTC market literature.

Most fixed-income markets are OTC, which likely affects firms' borrowing costs and their capital structure. He & Milbradt (2014) study the interplay between corporate bond liquidity and default by integrating an OTC secondary market into a capital-structure model à la Leland (1994). They show in particular that, when lower fundamentals push firms closer to default, the anticipation of post-default OTC market illiquidity depresses prices further, makes it more costly

to roll over, and ultimately accelerates default. Chen et al. (2018) and d'Avernas (2017) build on this framework to decompose yield spreads. Other papers sharing a similar corporate financing focus but exploring different channels include those by Hugonnier et al. (2015), Arseneau et al. (2017), Kozlowski (2018), Bethune et al. (2019), Cui & Radde (2019), and Roh (2019).

Monetary policy is implemented in the OTC market for federal funds. This has been studied in partial equilibrium by Afonso & Lagos (2015a,b), Bech & Monnet (2016), Armenter & Lester (2017), Afonso et al. (2019), Lagos & Navarro (2019), and Wong & Zhang (2019), and in general equilibrium by Bianchi & Bigio (2014) and Bigio & Sannikov (2019). In particular, in Bianchi & Bigio's (2014) paper, banks borrow extra reserves in an OTC interbank market to settle payment shocks. Ex ante, banks accumulate a buffer of reserve to reduce their reliance on OTC markets, which affects lending and aggregate economic activity. Malamud & Schrimpf (2017) develop a market setting in which some asset markets are organized as OTC markets with imperfectly competitive intermediaries, resulting in an imperfect monetary policy pass-through. Another approach to monetary policy is to explicitly consider that money is a means of payment for assets. While Lucas (1990) assumed a competitive asset market, a recent literature has considered OTC asset markets, with new insights into the impact of monetary policy. Research in this area includes work by Geromichalos & Herrenbrueck (2016), Geromichalos et al. (2016), Mattesini & Nosal (2016), Lagos & Zhang (2019a,b, 2020), and Lebeau (2019).

An emerging literature considers international economics applications. Geromichalos & Jung (2018), Malamud & Schrimpf (2018), and Bianchi et al. (2018) study the foreign exchange market, while Passadore & Xu (2018) and Chaumont (2018) study the interplay between liquidity and default for sovereign debt by integrating a secondary OTC market in the model of Eaton & Gersovitz (1981).

3.13. Calibration and Structural Estimation

Ever since Duffie et al.'s (2007) paper, many authors have combined their theoretical analyses with calibrations. The list includes Vayanos & Weill (2008) for the spread between on- and off-the-run bonds; He & Milbradt (2014), d'Avernas (2017), and Chen et al. (2018) for decompositions of credit spread between default and liquidity components; Passadore & Xu (2018) and Chaumont (2018) for sovereign spreads; Bianchi & Bigio (2014), Afonso & Lagos (2015a), Armenter & Lester (2017), and Afonso et al. (2019) for quantitative policy experiments in the federal funds market; Pagnotta & Philippon (2018) for a welfare analysis of speed competitions between exchanges; Hugonnier et al. (2020) for a decomposition of the gains from trade in the municipal bond market; and Kozlowski (2018) for a quantitative analysis of maturity choice when debt is traded in OTC markets.

Comparatively, the structural estimation of search models of OTC markets remains underdeveloped. Perhaps this is because it must overcome at least two significant obstacles: Search models must be rich enough to confront the data, and data must be sufficiently rich to confront implications that are unique to search models. Among structural studies of OTC market, Feldhütter (2012) identifies times of strong selling pressure in the corporate bond market; Gavazza (2016) studies the welfare impact of intermediaries in the secondary market for commercial aircraft; Brancaccio et al. (2017) study the contribution of experimentation to learn about order flow in the municipal bond market; Hendershott et al. (2020) study the formation of dealer-client relationships in the corporate bond market and the welfare impact of unbundling trade and nontrade services provided by intermediaries; and Liu (2020) studies the dynamic search process of dealers. Structural estimations of network models include work by Gofman (2014, 2017) and Eisfeldt et al. (2018).

4. CONCLUSION

Many asset markets have a decentralized OTC structure—for example, the markets for fixed-income securities, some derivatives, repos, and federal funds. Over the last two decades, OTC markets have been analyzed in a large theoretical and empirical literature. In this review, I have focused my attention on the branch of the literature that studies OTC markets through the lens of search-and-matching theory. I have developed a benchmark model to illustrate the key assumptions and economic forces at play in existing work, and I have discussed some of the key insights generated by the literature. Much work remains to be done, in particular to develop structural models that are sufficiently rich to be estimated based on detailed transaction-based microeconomic data.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This paper benefited from comments and suggestions by David Lindsay, Stephen Morris, Semih Üslü, Mengbo Zhang, and Diego Zúñiga. David Lindsay provided expert research assistance. I would like to thank Batchimeg Sambalaibat for sharing data about trading volume across markets.

LITERATURE CITED

Afonso G. 2010. Liquidity and congestion. *J. Financ. Int.* 20:324–60

Afonso G, Armenter R, Lester B. 2019. A model of the federal funds market: yesterday, today, and tomorrow. *Rev. Econ. Dyn.* 33:177–204

Afonso G, Kovner A, Schoar A. 2014. *Trading partner in the interbank market*. Staff Rep. No. 620, Fed. Reserve Bank, New York

Afonso G, Lagos R. 2011. *Trade dynamics in the federal funds market*. Work. Pap., Fed. Reserve Bank, New York

Afonso G, Lagos R. 2015a. The over-the-counter theory of the fed funds market: a primer. *J. Money Credit Bank.* 47:127–54

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

Akn Ş Nuray, Platt BC. 2019. *Transition dynamics in equilibrium search*. Work. Pap., Brigham Young Univ., Provo, UT

Amador M, Weill PO. 2006. *Learning by matching*. Work. Pap., Univ. Calif., Los Angeles

Amihud Y, Mendelson H. 1991. Liquidity, maturity, and the yield on U.S. treasury securities. *J. Finance* 46:479–86

Amihud Y, Mendelson H, Pedersen LH. 2005. Liquidity and asset prices. *Found. Trends Finance* 4:270–364

An Y. 2019. *Competing with inventory in dealership markets*. Work. Pap., John Hopkins Univ., Baltimore, MD

An Y, Zheng Z. 2018. *Conflicted immediacy provision*. Work. Pap., John Hopkins Univ., Baltimore, MD

Armenter R, Lester B. 2017. Excess reserves and monetary policy implementation. *Rev. Econ. Dyn.* 23:212–35

Arseneau DM, Rappoport DE, Vardoulakis AP. 2017. *Private and public liquidity provision in over-the-counter markets*. Work. Pap., Fed. Reserve Board, Washington, DC

Atkeson A, Eisfeldt A, Weill PO. 2013. *The market of OTC derivatives*. Work. Pap., Univ. Calif., Los Angeles

Atkeson A, Eisfeldt A, Weill PO. 2015. Entry and exit in OTC derivatives markets. *Econometrica* 83:2231–92

Aymanns C, Co-Pierre G, Golub B. 2018. *Illiquidity spirals in coupled over-the-counter markets*. Work. Pap., Harvard Univ., Cambridge, MA

Babus A. 2019. *Market for financial innovation*. Work. Pap., Washington Univ. St. Louis, St. Louis, MO

Babus A, Farboodi M. 2018. *The hidden cost of strategic opacity*. Work. Pap., Washington Univ. St. Louis, St. Louis, MO
Babus A, Hu TW. 2018. Endogenous intermediation in over-the-counter markets. *J. Financ. Econ.* 125:200–15
Babus A, Kondor P. 2018. Trading and information diffusion in OTC markets. *Econometrica* 86:1727–69
Babus A, Parlatore C. 2018. *Strategic fragmented markets*. Work. Pap., Washington Univ. St. Louis, St. Louis, MO
Bao J, O'Hara M, Zhou A. 2018. The Volcker rule and market making in times of stress. *J. Financ. Econ.* 130:95–113
Bech M, Monnet C. 2016. A search-based model of the interbank money market and monetary policy implementation. *J. Econ. Theory* 164:32–67
Bessembinder H, Jacobsen S, Maxwell W, Venkataraman K. 2018. Capital commitment and illiquidity in corporate bonds. *J. Finance* 73:1615–61
Bessembinder H, Maxwell W. 2008. Markets: transparency and the corporate bond market. *J. Econ. Perspect.* 22:217–34
Bethune Z, Sultanum B, Trachter N. 2018. *An information-based theory of financial intermediation*. Tech. Rep., Fed. Reserve Bank, Richmond, VA
Bethune Z, Sultanum B, Trachter N. 2019. Asset issuance in over-the-counter markets. *Rev. Econ. Dyn.* 33:4–29
Biais B. 1993. Price formation and equilibrium liquidity in fragmented and centralized markets. *Rev. Financ. Stud.* 48:157–85
Biais B, Green RC. 2019. The microstructure of the bond market in the 20th century. *Rev. Econ. Dyn.* 33:250–71
Biais B, Hombert J, Weill PO. 2014. Equilibrium pricing and trading volume under preference uncertainty. *Rev. Econ. Stud.* 81:1401–37
Biais B, Weill PO. 2009. *Liquidity shocks and order book dynamics*. Work. Pap., Toulouse Sch. Econ., Toulouse, Fr.
Bianchi J, Bigio S. 2014. *Banks, liquidity management, and monetary policy*. NBER Work. Pap. 20490
Bianchi J, Bigio S, Engel C. 2018. *Payments, liquidity, and exchange rates*. Work. Pap., Univ. Calif., Los Angeles
Bigio S, Sannikov Y. 2019. *A model of credit, money, interest, and price*. Work. Pap., Univ. Calif., Los Angeles
Blouin MR, Serrano R. 2001. A decentralized market with common values uncertainty: non-steady states. *Rev. Econ. Stud.* 68:323–46
Brancaccio G, Li D, Schurhoff N. 2017. *Learning by trading: the case of the U.S. market for municipal bond*. Work. Pap., Cornell Univ., Ithaca, NY
Burdett K, Judd KL. 1983. Equilibrium price dispersion. *Econometrica* 51:955–69
Camargo B, Lester B. 2014. Trading dynamics in decentralized markets with adverse selection. *J. Econ. Theory* 153:534–68
Cespa G, Vives X. 2018. *Exchange competition, entry, and welfare*. Work. Pap., City Univ. London, London
Chang B. 2018. Adverse selection and liquidity distorsion. *Rev. Econ. Stud.* 85:275–306
Chang B, Zhang S. 2018. *Endogenous market making and network formation*. Work. Pap., London Sch. Econ., London
Chaumont G. 2018. *Sovereign debt, default risk, and the liquidity of government bonds*. Work. Pap., Pa. State Univ., University Park
Chen H, Cui R, He Z, Milbradt K. 2018. Quantifying liquidity and default risks of corporate bonds over the business cycle. *Rev. Financ. Stud.* 31:852–97
Chiu J, Koeppl T. 2016. Trading dynamics with adverse selection and search: market freeze, intervention and recovery. *Rev. Econ. Stud.* 83:969–1000
Coles MG, Smith E. 1992. Marketplaces and matching. *Int. Econ. Rev.* 39:239–54
Colliard JE, Demange G. 2014. *Asset dissemination through dealer markets*. Work. Pap., HEC Paris, Jouy-en-Josas, Fr.
Colliard JE, Foucault T, Hoffmann P. 2018. *Inventory management, dealers' connections, and prices in OTC markets*. Work. Pap., HEC Paris, Jouy-en-Josas, Fr.
Cui W, Radde S. 2019. Search-based endogenous asset liquidity and the macroeconomy. *J. Eur. Econ. Assoc.* In press. **https://doi.org/10.1093/jeea/jvz037**

Cujean J, Praz R. 2013. *Asymmetric information and inventory concerns in over-the-counter markets*. Work. Pap., Univ. Bern, Bern, Switz.
d'Avernas A. 2017. *Disentangling credit spreads and equity volatility*. Work. Pap., Stockh. Sch. Econ., Stockholm, Swed.
Di Maggio M. 2013. *Market turmoil and destabilizing speculation*. Work. Pap., Mass. Inst. Technol., Cambridge
Diamond PA. 1971. A model of price adjustment. *J. Econ. Theory* 3:156–68
Diamond PA. 1982. Aggregate demand management in search equilibrium. *J. Political Econ.* 90:881–94
Dick-Nielsen J, Rossi M. 2019. The cost of immediacy for corporate bonds. *Rev. Financ. Stud.* 32:1–41
Duffie D. 2010. Presidential address: asset price dynamics with slow-moving capital. *J. Finance* 65:1237–67
Duffie D, Dworczak P, Zhu H. 2017. Benchmarks in search markets. *J. Finance* 72:1983–2044
Duffie D, Gârleanu N, Pedersen LH. 2002. Securities lending, shorting, and pricing. *J. Financ. Econ.* 66:307–39
Duffie D, Gârleanu N, Pedersen LH. 2005. Over-the-counter markets. *Econometrica* 73:1815–47
Duffie D, Gârleanu N, Pedersen LH. 2007. Valuation in over-the-counter markets. *Rev. Financ. Stud.* 20:1865–900
Duffie D, Malamud S, Manso G. 2009. Information percolation with equilibrium search dynamics. *Econometrica* 77:1513–74
Duffie D, Malamud S, Manso G. 2014. Information percolation in segmented markets. *J. Econ. Theory* 153:1–32
Duffie D, Manso G. 2007. Information percolation in large markets. *Am. Econ. Rev.* 97:203–9
Dugast J. 2018. Unscheduled news and market dynamics. *J. Finance* 78:2537–86
Dugast J, Üslü S, Weill PO. 2019. *A theory of participation in OTC and centralized markets*. Work. Pap., Paris Dauphine Univ., Paris
Eaton J, Gersovitz M. 1981. Debt with potential repudiation: theoretical and empirical analysis. *Rev. Econ. Stud.* 48:289–309
Eisfeldt A, Herskovic B, Rajan S, Siriwardane EN. 2018. *OTC intermediaries*. Work. Pap., Univ. Calif., Los Angeles
Farboodi M. 2014. *Intermediation and voluntary exposure to counterparty risk*. Work. Pap., Mass. Inst. Technol., Cambridge, MA
Farboodi M, Jarosch G, Menzio G, Wiriadinata U. 2018a. *Intermediation as rent extraction*. Tech. Rep., Mass. Inst. Technol., Cambridge
Farboodi M, Jarosch G, Shimer R. 2018b. *The emergence of market structure*. Work. Pap., Mass. Inst. Technol., Cambridge
Feldhütter P. 2012. The same bond at different prices: identifying search frictions and selling pressures. *Rev. Financ. Stud.* 25:1155–206
Fermanian JD, Guéant O, Pu J. 2016. The behavior of dealers and clients on the European corporate bond market: the case of multi-dealer-to-client platforms. arXiv:1511.07773 [q-fin.ST]
Gabrovski M, Kospentaris I. 2020. *Intermediation in over-the-counter markets with price transparency*. Work. Pap., Univ. Hawaii, Manoa
Gârleanu N. 2009. Portfolio choice and pricing in illiquid markets. *J. Econ. Theory* 144:532–64
Gavazza A. 2016. An empirical equilibrium model of a decentralized asset market. *Econometrica* 84:1755–98
Gehrig T. 1993. Intermediation in search markets. *J. Econ. Manag. Strategy* 2:97–120
Geromichalos A, Herrenbrueck L. 2016. Monetary policy, asset prices and liquidity in over-the-counter markets. *J. Money Credit Bank.* 48:35–79
Geromichalos A, Herrenbrueck L, Salyer K. 2016. A search-theoretic model of the term premium. *Theor. Econ.* 11:897–935
Geromichalos A, Jung KM. 2018. An over-the-counter approach to the FOREX market. *Int. Econ. Rev.* 59:859–905
Glebkin S, Shen J, Yueshen BZ. 2019. *Simultaneous multilateral search*. Work. Pap., INSEAD, Singapore
Glode V, Opp CC. 2020. Over-the-counter versus limit-order markets: the role of traders' expertise. *Rev. Financ. Stud.* 33:866–915
Gofman M. 2014. *A network-based analysis of over-the-counter markets*. Work. Pap., Univ. Rochester, Rochester, NY

Gofman M. 2017. Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *J. Financ. Econ.* 124:113–14
Golosov M, Lorenzoni G, Tsyvinski A. 2014. Decentralized trading with private information. *Econometrica* 82:1055–91
Guerrieri V, Shimer R. 2014. Dynamic adverse selection: a theory of illiquidity, fire sales, and flight to quality. *Am. Econ. Rev.* 104:1875–908
Guerrieri V, Shimer R. 2018. Markets with multidimensional private information. *Am. Econ. J. Microecon.* 85:1502–42
Guerrieri V, Shimer R, Wright R. 2010. Adverse selection in competitive search equilibrium. *Econometrica* 78:1823–62
Harris L. 2003. *Trading and Exchanges: Market Microstructure for Practitioners*. New York: Oxford Univ. Press
Hatheway F, Kwan A, Zheng H. 2017. An empirical analysis of market segmentation in U.S. equity markets. *J. Financ. Quant. Anal.* 52:2399–427
He Z, Milbradt K. 2014. Endogenous liquidity and defaultable bonds. *Econometrica* 82:1443–508
Hendershott T, Li D, Livdan D, Schürhoff N. 2020. Relationship trading in OTC markets. *J. Finance* 75:683–734
Hugonnier J. 2012. *Speculative behavior in decentralized markets*. Work. Pap., Swiss Finance Inst., Zürich, Switz.
Hugonnier J, Lester B, Weill PO. 2014. *Heterogeneity in decentralized asset markets*. NBER Work. Pap. 20746
Hugonnier J, Lester B, Weill PO. 2020. Frictional intermediation in over-the-counter markets. *Rev. Econ. Stud.* 87:1432–69
Hugonnier J, Malamud S, Morellec E. 2015. Capital supply uncertainty, cash holdings, and investment. *Rev. Financ. Stud.* 28:391–445
ISDA (Int. Swap Deriv. Assoc.). 2018. *Key trends in the size and composition of OTC derivatives markets*. Tech. Rep., Int. Swap Deriv. Assoc., New York
Kawakami K. 2017. Welfare consequences of information aggregation and optimal market size. *Am. Econ. J. Microecon.* 9:303–23
Kozlowski J. 2018. *Long-term finance and investment with frictional asset markets*. Work. Pap., Fed. Reserve Bank St. Louis, St. Louis, MO
Lagos R, Navarro G. 2019. *Monetary operating procedures in the fed funds market: theory and policy analysis*. Work. Pap., New York Univ., New York
Lagos R, Rocheteau G. 2007. Search in asset markets: market structure, liquidity, and welfare. *Am. Econ. Rev.* 97:198–202
Lagos R, Rocheteau G. 2009. Liquidity in asset markets with search frictions. *Econometrica* 77:403–26
Lagos R, Rocheteau G, Weill PO. 2011. Crises and liquidity in OTC markets. *J. Econ. Theory* 146:2169–205
Lagos R, Rocheteau G, Wright R. 2017. Liquidity: a new monetarist perspective. *J. Econ. Lit.* 55:371–440
Lagos R, Zhang S. 2019a. Monetary exchange in bilateral over-the-counter markets. *Rev. Econ. Dyn.* 33:205–27
Lagos R, Zhang S. 2019b. *The limit of monetary economics: on money as a medium of exchange in near-cashless economies*. Work. Pap., New York Univ., New York
Lagos R, Zhang S. 2020. Turnover liquidity and the transmission of monetary policy. *Am. Econ. Rev.* 110:1635–72
Lauermann S, Merzyn W, Gábor V. 2018. Learning and price discovery in a search model. *Rev. Econ. Stud.* 85:1159–92
Lebeau L. 2019. *Credit frictions and participation in OTC markets*. Work. Pap., Univ. Calif., Irvine
Lee T, Wang C. 2018. *Why trade over the counter? When investors want price discrimination*. Work. Pap., Cent. Eur. Univ., Budapest, Hung.
Leland HE. 1994. Corporate debt value, bond covenants, and optimal capital structure. *J. Finance* 49:1213–43
Lester B, Rocheteau G, Weill PO. 2015. Competing for order flow in OTC markets. *J. Money Credit Bank.* 47:77–126
Lester B, Shourideh A, Venkateswaran V, Zetlin-Jones A. 2018. *Market-making with search and information frictions*. Work. Pap., Fed. Reserve Bank, Philadelphia, PA
Lester B, Shourideh A, Venkateswaran V, Zetlin-Jones A. 2019. Screening and adverse selection in frictional markets. *J. Political Econ.* 127:338–77

Li D, Schürhoff N. 2019. Dealer networks. *J. Finance* 74:91–144
Li Q. 2018. *Securitization and liquidity creation in markets with adverse selection*. Work. Pap., Pa. State Univ., University Park
Li W, Song Z. 2019. *Dealers as information intermediaries in over-the-counter markets*. Work. Pap., John Hopkins Univ., Baltimore, MD
Liu S. 2020. Dealer's search intensity in U.S. corporate bond markets. Work. Pap., Univ. Calif., Los Angeles
Liu Y, Vogel S, Zhang Y. 2018. *Electronic trading in OTC markets versus centralized exchange*. Res. Pap. 18-19, Swiss Finance Inst., Zürich, Switz.
Lucas RE Jr. 1990. Liquidity and interest rates. *J. Econ. Theory* 50:237–64
Lucas RE Jr. 2012 (1989). The effects of monetary shocks when prices are set in advance. In *Collected Papers on Monetary Economy*, ed. RE Lucas Jr., M Gilman, pp. 272–99. Cambridge, MA: Harvard Univ. Press
Malamud S, Rostek M. 2017. Decentralized exchange. *Am. Econ. Rev.* 107:3320–62
Malamud S, Schrimpf A. 2017. *Intermediation markups and monetary policy pass-through*. Res. Pap. 16-75, Swiss Finance Inst., Zürich, Switz.
Malamud S, Schrimpf A. 2018. *An intermediation-based model of exchange rates*. Res. Pap. 18-14, Swiss Finance Inst., Zürich, Switz.
Manea M. 2018. Intermediation and resale in networks. *J. Political Econ.* 126:1250–301
Mattesini F, Nosal E. 2016. Liquidity and asset prices in a monetary model with OTC asset markets. *J. Econ. Theory* 164:187–217
Maurin V. 2018. *Liquidity fluctuations in over-the-counter markets*. Work. Pap., Stockh. Sch. Econ., Stockholm, Swed.
Miao J. 2005. A search model of centralized and decentralized trade. *Rev. Econ. Dyn.* 9:68–92
Milbradt K. 2017. *Asset heterogeneity in over-the-counter markets*. Work. Pap., Kellogg Sch. Manag., Evanston, IL
Mortensen DT, Pissarides CA. 1994. Job creation and job destruction in the theory of unemployment. *Rev. Econ. Stud.* 61:397–415
Neklyudov A. 2019. Bid-ask spreads and the over-the-counter interdealer markets: core and peripheral dealers. *Rev. Econ. Dyn.* 33:57–84
Newman YS, Rierson MA. 2003. *Illiquidity spillovers: theory and evidence from European telecom bond issuance*. Job Mark. Pap., Stanford Univ., Stanford, CA
Nosal E, Rocheteau G. 2011. *Money, Payments, and Liquidity*. Cambridge, MA: MIT Press
Nosal E, Wong YY, Wright R. 2019. *Intermediation in markets for goods and markets for assets*. Work. Pap., Fed. Reserve Bank, Atlanta, GA
Pagano M. 1989. Trading volume and asset liquidity. *Q. J. Econ.* 104:255–74
Pagnotta ES, Philippon T. 2018. Competing on speed. *Econometrica* 86:1067–115
Passadore J, Xu Y. 2018. *Illiquidity in sovereign debt markets*. Work. Pap., Einaudi Inst. Econ. Finance, Rome, Italy
Petrongolo B, Pissarides CA. 2001. Looking into the black box: a survey of the matching function. *J. Econ. Lit.* 39:390–431
Praz R. 2015. *Equilibrium asset pricing with both liquid and illiquid markets*. Work. Pap., Cph. Bus. Sch., Frederiksberg, Den.
Riggs L, Onur E, Reiffen D, Zhu H. 2018. *Swap trading after Dodd-Frank: evidence from index CDS*. Work. Pap., Commod. Futures Trading Comm., Washington, DC
Rocheteau G, Weill PO. 2011. Liquidity in frictional asset markets. *J. Money Credit Bank.* 43:261–82
Roh HS. 2019. *Repo specialness in the transmission of quantitative easing*. Work. Pap., Stanford Univ., Stanford, CA
Rosu I. 2009. A dynamic model of the limit-order book. *Rev. Financ. Stud.* 22:4601–41
Rubinstein A. 1982. Perfect equilibrium in a bargaining model. *Econometrica* 50:97–109
Rust J, Hall G. 2003. Middlemen versus market makers: a theory of competitive exchange. *J. Political Econ.* 111:353–403
Sambalaibat B. 2015. *A theory of liquidity spillover between bond and CDS markets*. Work. Pap., Indiana Univ., Bloomington

Sambalaibat B. 2018. *Endogenous specialization in dealer networks*. Work. Pap., Indiana Univ., Bloomington
Shen J, Wei B, Yan H. 2015. *Financial intermediation chains in an OTC market*. Work. Pap., Fed. Reserve Bank, Atlanta, GA
SIFMA (Secur. Ind. Financ. Mark. Assoc.). 2018. *U.S. fixed income market structure primer.* Tech. Rep., Secur. Ind. Financ. Mark. Assoc., New York
Sultanum B. 2018. Financial fragility and OTC markets. *J. Econ. Theory* 177:618–58
Trejos A, Wright R. 2016. Search-based models of money and finance: an integrated approach. *J. Econ. Theory* 164:10–31
Tse CY, Xu Y. 2018. *Inter-dealer trades in OTC markets—Who buys and who sells?* Work. Pap., Univ. Hong-Kong
Tuttle L. 2014. *OTC trading: description of non-ATS OTC trading in national market system stocks*. Tech. Rep., US Secur. Exch. Comm., Washington, DC
Üslü S. 2019. Pricing and liquidity in decentralized asset markets. *Econometrica* 87:2079–140
Üslü S, Velioğlu G. 2019. *Liquidity in the cross section of OTC assets*. Work. Pap., John Hopkins Univ., Baltimore, MD
Vayanos D, Wang T. 2007. Search and endogenous concentration of liquidity in asset markets. *J. Econ. Theory* 136:66–104
Vayanos D, Weill PO. 2008. A search-based theory of the on-the-run phenomenon. *J. Finance* 63:1361–98
Vogel S. 2019. *When to introduce electronic trading platforms in over-the-counter markets?* Work. Pap. Swiss Finance Inst., Zürich, Switz.
Wang C. 2018. *Core-periphery trading networks*. Work. Pap., Univ. Pa., Philadelphia
Weill PO. 2004. Essays on liquidity in financial markets. PhD Thesis, Stanford Univ., Stanford, CA
Weill PO. 2007. Leaning against the wind. *Rev. Econ. Stud.* 74:1329–54
Weill PO. 2008. Liquidity premia in dynamic bargaining markets. *J. Econ. Theory* 140:66–96
Weill PO. 2011. Liquidity provision in capacity constrained markets. *Macroecon. Dyn.* 15:119–44
Williams B. 2014. *Search, liquidity, and retention: screening multidimensional private information*. Work. Pap., New York Univ., New York
Wolinsky A. 1990. Information revelation in a market with pairwise meetings. *Econometrica* 58:1–23
Wong R, Zhang M. 2019. *Disintermediating the federal funds market*. Work. Pap., Univ. Calif., Los Angeles
Wright R, Kircher P, Guerrieri V, Julien B. 2017. *Directed search: a guided tour*. NBER Work. Pap. 23884
Yang M, Zeng Y. 2019. *The coordination of intermediation*. Work. Pap., Duke Univ., Durham, NC
Yavaş A. 1992. Marketmakers versus matchmakers. *J. Financ. Int.* 2:33–58
Yoon JH. 2017. *Endogenous market structure: over-the-counter versus exchange trading*. Job Mark. Pap., Univ. Wisconsin-Madison, Madison
Zhang S. 2017. Liquidity missallocation in an over-the-counter market. *J. Econ. Theory* 174:16–56
Zhu H. 2011. Finding a good price in opaque over-the-counter markets. *Rev. Financ. Stud.* 25:1255–85
Zou J. 2019. *Information acquisition and liquidity traps in over-the-counter markets*. Work. Pap., INSEAD, Singapore