Endogenous Specialization and Dealer Networks*

Artem Neklyudov[†] and Batchimeg Sambalaibat[‡]

October 2016

Preliminary and Incomplete
The latest version: www.dropbox.com/s/szjrkl602c3ra3t/Clientele.pdf?dl=0

OTC markets exhibit a core-periphery network: 10-30 central dealers trade frequently and with many dealers, while hundreds of peripheral dealers trade sparsely and with few dealers. Existing work rationalize this phenomenon with exogenous dealer heterogeneity. We build a search-based model of network formation and propose that a core-periphery network arises from specialization. Dealers endogenously specialize in different clients with different liquidity needs. The clientele difference across dealers, in turn, generates dealer heterogeneity and the core-periphery network: The dealers specializing in clients who trade frequently form the core, while the dealers specializing in buy-and-hold investors form the periphery.

Keywords: Network formation, core-periphery, clientele effect, specialization, intermediation chains, over-the-counter markets, search frictions.

---

*We thank Barney Hartman-Glaser, Burton Hollifield, Norman Schuerhoff, Chris Telmer, Pierre-Olivier Weill, Ariel Zetlin-Jones, seminar participants at UCLA Anderson, the University of Wisconsin, Indiana University, Stockholm School of Economics, Copenhagen Business School, European University Institute, conference participants and discussants at the ASU Sonoran Winter Finance Conference, Society of Economic Dynamics, FIRS, the 4th Economic Networks and Finance Conference, McGill University Risk Management Conference, HEC-McGill Winter Finance Workshop, Minnesota Mini-Finance Conference, 9th Swiss Conference on Financial Intermediation, Southwest Search and Matching Workshop, Federal Reserve Bank of St. Louis Summer Workshop on Money, Banking, Payments, and Finance, Women in Microstructure Workshop, and the Annual Central Bank Workshop on the Microstructure of Financial Markets. We completed a part of this work while Batchimeg was a visiting scholar at the University of Lausanne. Batchimeg greatly acknowledges the University of Lausanne and SFI for the hospitality and financial support.

[†]University of Lausanne and SFI (artem.neklyudov@unil.ch)

[‡]Kelley School of Business, Indiana University (bsambala@iu.edu)

In over-the-counter (OTC) markets, transactions between dealers exhibit a core-periphery network. Ten to thirty highly interconnected dealers account for a majority of both dealer-to-dealer and client-to-dealer transactions. These dealers form the core, while hundreds of sparsely connected dealers trade infrequently and form the periphery. This network structure is not a one-time random event but is highly persistent over time. In particular, both the dealers' relative importance in the network and who they trade with are highly persistent.[1] Li and Schürhoff (2014) (LS hereon) document these patterns for the municipal bond market and Neklyudov, Hollifield, and Spatt (2014) (NHS hereon) for the asset-backed securities market.[2]

These stylized facts challenge existing models. Recent papers rationalize the core-periphery phenomenon with ex-ante dealer heterogeneity.[3] Current network models are one-time static models and hence cannot speak to the observed network persistence. Search models—a prominent class of models capturing OTC markets—imply that trading networks are random.

Thus, we still need to explain: How does dealer heterogeneity arise in the first place? And why do core and peripheral dealers co-exist? Any convincing explanation has to—at the same time—explain the observed network persistence. How do core dealers maintain their size and market share and persistently remain in the core?

We build a search-based model of network formation and show that dealer heterogeneity and the core-periphery network arise from specialization. Some dealers form the core because they specialize in investors who trade frequently (e.g. index funds). Because they cater to customers who trade frequently, core dealers receive a large volume of client orders. Their client orders, in turn, support the large volume of interdealer trades they transact and hence their centrality in the network. Conversely, the dealers that specialize in buy-and-hold investors (e.g. pension funds) form the periphery. Thus, how

[1] LS document the persistence in two dimensions. First, the probability that a top-ten central dealer remains a top-ten dealer month-to-month is 93%. The persistence is 97% for peripheral dealers. Second, if two dealers trade one month, the probability that they trade again the following month is 65%. In a random network, this probability is 1.4%.

[2] Other OTC markets exhibit similar networks. Afonso, Kovner, and Schoar (2013) and Bech and Atalay (2010), for example, document a core-periphery structure in the inter-bank lending market.

[3] In Atkeson, Eisfeldt, and Weill (2014), for example, the dealers with a larger number of traders form the core. In Zhong (2014) and Neklyudov (2012), the dealers with exogenously larger inventory capacity and superior trading technology, respectively, form the core. Hugonnier, Lester, and Weill (2014) and Chang and Zhang (2015) assume a heterogeneity in agents' preference for an asset. In the former, agents have idiosyncratic realizations of asset valuations; in the latter, agents have both heterogeneous volatility and idiosyncratic realizations. Recent network models fix agents' network centrality (see, for example, Gofman (2011), Kondor and Babus (2013), and Malamud and Rostek (2014)).

clients form around dealers determines the shape of the interdealer network. This insight is the main contribution of the paper.

We formalize this insight with a model that builds on Duffie, Garleanu, and Pedersen (2005) and, in particular, on Vayanos and Wang (2007). We add to their environment dealers and interdealer trades. Dealers are ex-ante identical, but customers have heterogenous liquidity needs. Some customers just buy and hold an asset; others buy knowing they will turn around and sell quickly. Dealers intermediate directly between customers, but also connect with other dealers to supplement their liquidity provision to customers. We assume a fully connected dealer network, but network weights (in particular, the transaction volumes between pairs of dealers) are endogenous.

In this environment, we show that both symmetric and asymmetric equilibria exist. The symmetric equilibrium features a circular network, where dealers have identical network centrality. This shows that client heterogeneity alone does not guarantee dealer heterogeneity. The asymmetric equilibrium, on the other hand, features a core-periphery network due to specialization and the heterogeneity that it creates.

In the asymmetric equilibrium, the endogenous dealer specialization works as follows. Clients tradeoff a dealer's ask-price versus its return service. Some dealers charge a high ask-price but, in return, offer a better service if the client has to return the bond: The dealer either buys back at a higher bid-price, executes the order more quickly, or both. Others charge a cheaper ask-price price but offer a worse return service. Buyers who expect to reverse their position quickly care more about what happens to them as a seller. They, as a result, choose the dealer based on its return service and are willing to pay the higher ask-price. Buy-and-hold investors, less concerned with turning into a seller later on, instead, choose the dealer offering the cheapest price. Thus, investors with different liquidity needs endogenously sort across different dealers. The clientele difference across dealers, in turn, supports the different prices and liquidity across dealers. It also generates, as previously explained, the heterogeneity in the volume of client orders, the volume of interdealer trades, and hence the network centrality across dealers.

Our second contribution lies in capturing the observed network persistence. The observed persistence challenges two central assumptions of search models. First, search models assume that agents' private valuations of an asset change randomly (as a way to generate trade in equilibrium). The assumption implies that agents' intermediation roles are random.[4] Second, the

[4]That is, Goldman Sachs, a core dealer, can randomly become a mom-and-pop peripheral asset management firm one period and then randomly switch back to being Goldman Sachs another period. In Hugonnier, Lester, and Weill (2014), for example, agents with

standard models assume that agents trade through random search and match and thus abstract from repeated trades between agents. We relax both of these assumptions. We model clients and dealers separately and model valuation changes occurring with clients. Dealers' identities and their equilibrium roles (e.g. whether they are a core or peripheral), as a result, remain stable and hence the persistence in the intermediation roles. The stability of dealer identities allows us to model explicit network links between dealers. Dealers, as a result, trade with each other repeatedly and hence the persistence in the interdealer trades.[5]

Additionally, we show that core and peripheral dealers play the following roles. On the interdealer market, core dealers supply liquidity (by volume and execution speed) to other dealers but charge wide bid-ask spreads. Peripheral dealers consume that liquidity and pass it down to their clients (specifically, the execution speed and wide bid-ask spreads). They rely more on the interdealer market and on long intermediation chains for their liquidity service to clients. Bonds, as a result, cycle through the economy starting with core dealers' clients, then the interdealer network, and eventually end with buy-and-hold investors, who are concentrated with peripheral dealers. The cycle repeats when a buy-and-hold investor experiences a liquidity shock and sells the bond. The sell order, in turn, primarily gets absorbed via the interdealer network by core dealers and their clients. Thus, core dealers serve as a central conduit in transmitting assets through the economy from one end-customer to another.

Finally, we highlight three additional results. First, we show that specialization and the resulting core-periphery network are socially desirable and dominate a circular network. Second, interconnectedness among dealers improves bond market liquidity: It increases the aggregate volume of transactions, narrows bid-ask spreads, and speeds up transaction times. Greater liquidity, in turn, alleviates misallocations and improves both the customer welfare and dealer profits. Third, market fragmentation (captured by the aggregate number of dealers) also increases the total welfare. Whether the increase in the welfare accrues to clients or dealers, however, depends on their relative bargaining powers.

We proceed as follows. Section 1 presents the model. In Section 2, we derive the asymmetric specialization equilibrium and show that the dealer

an intermediate asset valuation resemble core dealers, while agents with extreme valuations resemble peripheral dealers. As agents randomly switch between different valuations, a dealer that is a core dealer one period may randomly become a peripheral dealer the next period and vice versa. Similarly, in Shen, Wei, and Yan (2015), an agent randomly switches between trading like a dealer versus like a client.

[5] Also, clients in our model choose dealers and trade repeatedly with their dealers.

network has a core-periphery structure. Section 3 compares liquidity and prices that core and peripheral dealers provide to customers and, on the interdealer market, to other dealers. Section 4 derives additional results on dealer interconnectedness, market fragmentation, and welfare. In Section 5, we discuss our assumptions. Section 6 concludes.

Related Literature

We close the gap between the network and search literatures: We provide a novel way to think about dealers and dealer networks in an environment with search and matching frictions. We depart from Duffie, Garleanu, and Pedersen (2005) (DGP) in an important way: From the perspective of clients, dealers are segmented. In DGP, end-customers trade with one another directly through random search and match, but also frictionlessly with any dealer. Thus, the implicit assumption in DGP is a zero cost of forming a client-dealer relationship. In contrast, our model features dealer segmentation and thus implicitly assumes a fixed cost of forming a relationship with a dealer. This simple tweak (dealer segmentation) allows us to model and study (1) clients' endogenous choice over dealers, (2) multiple dealers, (3) the intermediation chain among dealers, and (4) dealer heterogeneity.[6]

Our paper relates to recent models with implications on trading networks among agents. In Atkeson, Eisfeldt, and Weill (2014), for example, the dealer banks with a larger number of traders and intermediate exposures to aggregate risk resemble a core dealer. In Zhong (2014) and Neklyudov (2012), the dealers with an exogenously larger inventory capacity and a superior trading technology, respectively, form the core. In Hugonnier, Lester, and Weill (2014) and Shen, Wei, and Yan (2015), agents have idiosyncratic realizations of private valuations for an asset, and those with intermediate valuations intermediate the most and resemble a core dealer. In Chang and Zhang (2015), agents have both heterogeneous volatility and idiosyncratic realizations. In contrast to these papers, in our model, the heterogeneity across dealers arises endogenously.

In the network literature, a large strand studies networks in the interbank lending market (see, for example, Farboodi (2014) and Wang (2014)). We instead develop a model with a broader application to any OTC market. The model, as a result, predicts transaction volumes, bid-ask spreads, and liquidity provision. Other network models, such as Kondor and Babus (2013),

[6]For other search models applied to financial markets see, for example, Weill (2008), Vayanos and Weill (2008), Lagos and Rocheteau (2009), Duffie, Malamud, and Manso (2009), and Sambalaibat (2014).

are based on asymmetric information. In contrast, we offer a search-based network model. Yet another strand takes the network structure and hence the heterogeneity in network centrality as given (see, for example, Gofman (2011), Kondor and Babus (2013), and Malamud and Rostek (2014)). We allow for endogenous network weights.[7]

In our model, some dealers in equilibrium intermediate more dealer-to-dealer trades than other dealers. Bonds also travel through longer intermediation chains with peripheral dealers than with core dealers. Thus, our paper relates to models of intermediation chains (e.g., Viswanathan and Wang (2004), Glode and Opp (2014), Gofman (2011), Colliard and Demange (2014), Hugonnier, Lester, and Weill (2014), and Shen, Wei, and Yan (2015)).

1 Model

Time is continuous and goes from zero to infinity. There is one asset—a bond with supply S paying a coupon flow δ—and two sets of agents: customers and three ex-ante identical dealers. Dealers are indexed by $i \in N$, where $N = \{1, 2, 3\}$ is the set of dealers.[8] Everyone is risk neutral, infinitely lived, and discounts the future at a constant rate $r > 0$.

Customers

Customers are the end-users of the bond. As in standard search models, they have an idiosyncratic high or low valuation for the bond. High types derive a flow utility δ from holding the bond, while low types derive $\delta - x$, where $x > 0$ represents a disutility of holding the bond. High types thus in equilibrium want to own the bond; low types do not. Categorizing agents by their valuation and asset holding, we label them according to their equilibrium trading strategy: a buyer, owner, and seller.

Investors' valuations, moreover, change randomly, thus generating a need to rebalance their asset position and trade. In particular, high types experience a liquidity shock with intensity k and switch to a low type. The low

[7]The dealer network in our model is part exogenous and part endogenous. It is exogenous in that we assume a fully connected dealer network and that dealers do not choose who to link to. Thus, we implicitly assume a zero cost of forming a link. It is endogenous in that, once linked, link strengths (that is, network weights) are endogenous. Farboodi (2014) and Chang and Zhang (2015), for example, treat more formally the network formation process.

[8]Results on endogenous dealer specialization, which we show in the next section, hold for any number of dealers: $N \geq 2$. We need, however, at least $N \geq 3$ to derive the core-periphery results because with just two dealers, the amount of interdealer trades (and hence the network centrality) are necessarily the same across the two dealers.

state is an absorping state (that is, they do not switch back to a high type). Upon a liquidity shock, as a result, investors exit the economy, or if they own bonds, they first sell and then exit. Replenishing the exiting investors, new investors enter the economy as high type non-owners (that is, as buyers).

Investors, in addition, differ by their liquidity type, k: the rate with which they experience the liquidity shock. The distribution over k is given by the density function $\hat{f}(k)$ on support $[\underline{k}, \overline{k}]$.[9] A k-type investor expects to hold the bond for a period of $\frac{1}{k}$; thus, different liquidity types have different expected trading horizons. Those with a high switching rate (k) have a short trading horizon ($\frac{1}{k}$) and expect to have to sell quickly, while those with a small k expect to hold the bond longer. We refer to the former as liquidity investors and to the latter as buy-and-hold investors.

Investors can only buy and sell through one of the dealers. Upon entering the economy, a k-type buyer chooses dealer i with probability $\nu_i(k)$ according to

$$\nu_i(k) = \begin{cases} 1 & V_i^b(k) > \max_{j \neq i} V_j^b(k) \\ [0,1] \text{ if} & V_i^b(k) = \max_{j \neq i} V_j^b(k) \\ 0 & V_i^b(k) < \max_{j \neq i} V_j^b(k), \end{cases} \tag{1}$$

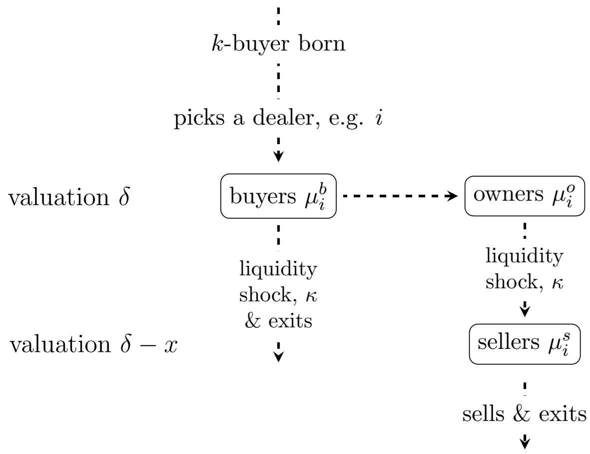
where $V_i^b(k)$ denotes the expected utility of a k-type buyer who is a customer of dealer i, and $\sum_{i \in N} \nu_i(k) = 1$. Once an investor chooses a dealer, we assume that, from then on, she can trade only through that dealer.

Figure 1 summarizes the life-cycle of investors. An investor enters the economy as a high type non-owner (i.e. as buyers), picks, say, dealer i, and becomes a buyer-client of that dealer. Upon buying the bond, she becomes an owner-client of the dealer. As an owner, she holds the bond until she experiences a liquidity shock and becomes a seller. Upon selling the bond, the investor exits the economy.

[9]The flow of new-entrants with liquidity types in $[k, k+dk]$ is $\hat{f}(k)dk$. We assume $\hat{f}(k)$ is a continuous strictly positive function.

Figure 1: Clients of Dealer i: Buyers, Owners, and Sellers

The figure illustrates in dashed (black) lines clients' life-cycle from a buyer, to an owner, to a seller. Upon a liquidity shock, an investor's bond valuation changes from δ to $\delta - x$, where x is a disutility of holding the bond.



We denote by μ_i^s, μ_i^b, and μ_i^o the total measure of sellers, buyers, and owners of dealer i, where

$$\mu_i^b \equiv \int_{\underline{k}}^{\overline{k}} \hat{\mu}_i^b(k)dk \tag{2}$$

$$\mu_i^o \equiv \int_{\underline{k}}^{\overline{k}} \hat{\mu}_i^o(k)dk. \tag{3}$$

The functions $\hat{\mu}_i^b(k)$ and $\hat{\mu}_i^o(k)$ are such that $\hat{\mu}_i^b(k)dk$ and $\hat{\mu}_i^o(k)dk$ are the measures of buyers and owners with switching rates k in $[k, k + dk]$.

Dealers and Intermediations

Dealers intermediate bonds for customers. They do so in two ways. First, a dealer pairs up buyers and sellers within its own client base according to

$$M_i^D \equiv \lambda_{\text{D}} \mu_i^s \mu_i^b, \tag{4}$$

where λ_{D} is an exogenous matching efficiency of a dealer.[10] Adopting the notation from LS and NHS, M_i^D is the volume of CDC (Client-Dealer-Client) intermediation chains, where the first C is the end-seller client, and the last

[10] A general functional form for the matching functions would be $M(\mu_b, \mu_s) = \lambda (\mu_b)^{\alpha_b} (\mu_s)^{\alpha_s}$. Thus, we implicitly assume: $\alpha_s = \alpha_b = 1$. Although constant returns to scale is standard in search models applied to labor markets, in the context of OTC financial markets, the standard assumption is increasing returns to scale. Weill (2008) shows that comparative statics from a model with increasing returns to scale fit better the stylized facts regarding, for example, liquidity and asset supply.

C is the end-buyer client. We assume dealers do not hold inventory: They buy from one client and instantly sell to another.

Second, a dealer intermediates for its clients by connecting with other dealers. We denote the set of dealer connections of dealer i with N_i and assume that each dealer is connected to every other dealer: $N_i = \{j \in N : j \neq i\}$ for all i. We define dealers i and j as connected if they share their clients with each other. In particular, using i's sellers and j's buyers, dealers i and j together produce $\lambda_{\text{DD}}\mu_i^s\mu_j^b$ matches (i.e. CDDC chains), where i is the *first* D in the chain, and λ_{DD} is a joint matching efficiency of the two dealers.[11] Analogously, using j's sellers and i's buyers, they produce $\lambda_{\text{DD}}\mu_j^s\mu_i^b$ CDDC chains, where i is now the *second* D in the chain. Summing across all dealers j that dealer i is connected to, the total volume of CDDC chains that dealer i intermediates is:

$$M_i^{DD} \equiv \underbrace{\lambda_{\text{DD}}\mu_i^s\left(\sum_{j\in N_i}\mu_j^b\right)}_{\text{C}\mathbf{D}\text{DC}} + \underbrace{\lambda_{\text{DD}}\left(\sum_{j\in N_i}\mu_j^s\right)\mu_i^b}_{\text{CD}\mathbf{D}\text{C}}. \tag{5}$$

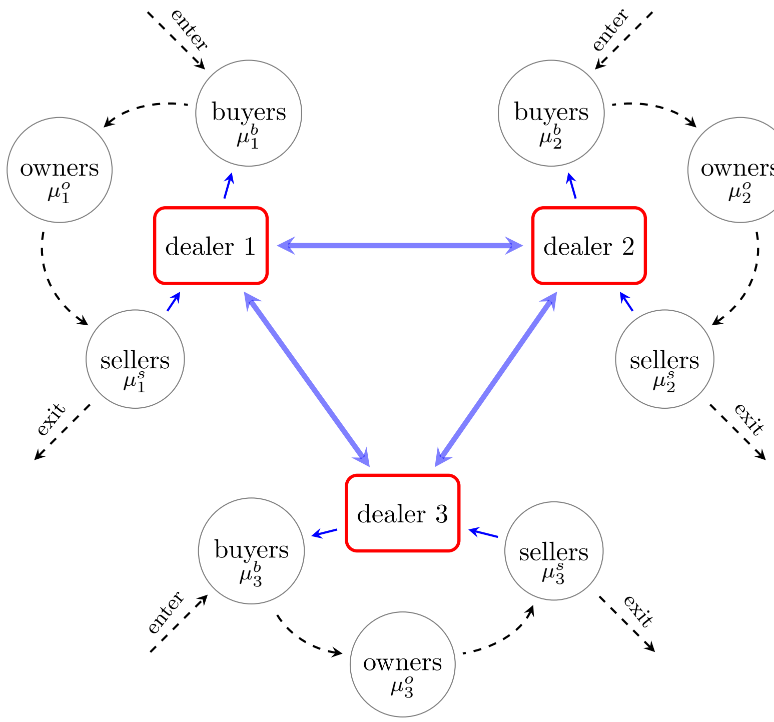
Comparing (5) with (4), if, for example, $\lambda_{\text{DD}} > \lambda_{\text{D}}$, two-dealer intermediation chains are more efficient than one-dealer chains. Figure 2 illustrates the environment.

In our environment, the source of inefficiency is that—due to matching frictions—investors with a low valuation for a bond (i.e. sellers) are stuck holding the bond despite the availability of willing buyers. Specifically, after receiving orders, dealers take time in producing matches and thereby create wait times for clients eventhough clients can instantly contact and submit an order with a dealer. Thus, trading frictions manifest as waiting periods after a client submits an order with a dealer. In a frictionless environment ($\lambda_{\text{D}} \to \infty$, $\lambda_{\text{DD}} \to \infty$), a customer would sell instantly, via their dealer, to another end-customer with a higher valuation (i.e. a buyer). Our specification is realistic. In practice, customers (as well as dealers themselves) can easily call up and put an order with a dealer, but immediate transactions are not guaranteed.

[11] CDDC means Client-Dealer-Dealer-Client chain, where the ordering captures the direction of the bond flow. The first C is the end-seller client, and the last C is the end-buyer client. The first D is the dealer buying from the end-seller and selling to the second dealer, and the second D is the dealer buying from the first D and selling to the end-buyer client.

Figure 2: Clients, Dealers, and Interdealer Trades

The figure illustrates the model environment. Dashed (black) lines represent clients' life-cycle between different client types (buyer, owner, and seller). Solid (blue) lines represent bond transaction flows. The sizes of circles represent the sizes of client measures.



Market Clearing

The supply of bonds circulating among customers of dealer i—denoted by s_i and endogenously determined—equals the measure of customers who currently hold the bond:

$$\int_{\underline{k}}^{\overline{k}} \hat{\mu}_i^o(k)dk + \mu_i^s = s_i. \tag{6}$$

For market clearing, the number of bonds circulating across all dealers' clients has to equal the aggregate supply of the bond, S:

$$\sum_{i \in N} s_i = S. \tag{7}$$

Interdealer Trades

We ensure that, in the steady state, a dealer is not growing or shrinking:

$$\lambda_{\text{DD}} \mu_i^s \left(\sum_{j \in N_i} \mu_j^b \right) = \lambda_{\text{DD}} \left(\sum_{j \in N_i} \mu_j^s \right) \mu_i^b. \tag{8}$$

The left- and right-hand sides are the total volume of bonds dealer i sells and buys on the interdealer market, respectively. Equating the two ensures that the dealer is neither a net buyer or a seller on the interdealer market.

Client Masses and Transitions

Customer masses have to be constant in the steady state. In particular, the flow of investors switching to a particular type has to equal the flow of investors switching out of that type. The mass of k-type buyers, as a result, is determined by

$$\overbrace{\hat{f}(k)\nu_i(k)dk}^{\text{inflow}} = \overbrace{k\hat{\mu}_i^b(k)dk + (\sum_{j\in N}\lambda_{ij}\mu_j^s)\hat{\mu}_i^b(k)dk}^{\text{outflow}}, \tag{9}$$

where $\lambda_{ij} = \lambda_{\mathrm{DD}}$ if $i \neq j$; otherwise, $\lambda_{ij} = \lambda_{\mathrm{D}}$. The left-hand side is the flow of type $k \in [k, k + dk]$ investors who become a buyer of dealer i. On the right-hand side, the first term is the flow of k-type buyers who experience a liquidity shock and exit the economy. The second term is the flow of buyers who get matched; in particular, buyers find a bond through their dealer with intensity $\sum_{j\in N}\lambda_{ij}\mu_j^s$. Appendix A analogously characterizes the owner and seller masses.

Prices

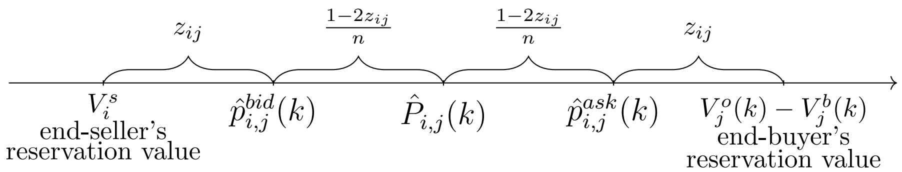
Prices arise from a sharing rule and are illustrated in Figure 3. Denoting by V_i^s, $V_i^b(k)$, and $V_i^o(k)$ the expected utility of a seller-, buyer-, and owner-client of dealer i, the reservation values of a buyer and a seller are $V_i^o(k) - V_i^b(k)$ and V_i^s, respectively. The total gains from trade is the difference between the buyer and seller's reservation values.

Prices are such that the end-seller of dealer i and the end-buyer of dealer j each capture z_{ij} fraction of the total gains from trade, where $z_{ij} = z_{\mathrm{DD}}$ if $i \neq j$ (i.e. 2-dealer chain); otherwise, $z_{ij} = z_{\mathrm{D}}$. We interpret z_{ij} as customers' bargaining power. Dealers split equally the remaining $1 - 2z_{ij}$ fraction. Prices, as a result, are a weighted average of buyer and sellers' reservation values. A seller-client of dealer i sells to his dealer at the bid price $\hat{p}_{i,j}^{bid}(k)$ given in (A.34), who turns around and sells to dealer j at the interdealer price $\hat{P}_{i,j}(k)$ in (A.35). Dealer j, in turn, sells to its buyer-client at the ask price $\hat{p}_{i,j}^{ask}(k)$ in (A.36). Prices are, thus, specific to the dealers and the end-customers involved in a chain.

When choosing dealers, however, a k-type buyer considers prices across all possible end-sellers that he could be matched with and, as a result, considers the expected ask-price of a dealer, $\overline{p}_i^{ask}(k)$. Similarly, a seller client considers the average bid-price across possible end-buyers she could be matched with, $\overline{p}_i^{bid}$. Appendix A characterizes these expected prices and the expected bid-ask spread $\overline{\phi}_i(k)$ customers face from their dealers. Equations (16)-(17) characterize the probability of getting matched with a buyer, m_i^b, and a seller, m_i^s, respectively.

Figure 3: Transaction Prices

The total gains from trade is the difference between the end-buyer and end-seller's reservation values. Prices, characterized in (A.34)-(A.36), are such that the two end-customers each capture z_{ij} fraction of the total surplus; dealers split equally the remaining $1-2z_{ij}$ fraction. The number of dealers involved in a chain is n.



Value Functions

Clients' value functions solve their optimization problem. Consider, for example, a k-type buyer who is a customer of dealer i. In a small time interval $[t+dt]$, a buyer could (a) receive a liquidity shock and exit the economy before he could purchase the bond (with probability kdt and get utility 0), (b) buy a bond (with probability $\sum_{j\in N}\lambda_{ij}\mu_j^s dt$ and get $V_i^o(k)-\hat{p}_{j,i}^{ask}(k)$), or (c) remain a buyer:

$$V_i^b(k) = (1-rdt)\Bigg(kdt0 + \sum_{j\in N}\lambda_{ij}\mu_j^s dt(V_i^o(k)-\hat{p}_{j,i}^{ask}(k)) + \\ +[1-kdt-\sum_{j\in N}\lambda_{ij}\mu_j^s dt]V_i^b(k)\Bigg). \tag{10}$$

Appendix A analogously derives the value functions of owner and seller types.

Our analysis focuses on the steady state equilibrium:

Definition. *A steady state equilibrium is expected utilities* $\{V_i^o(k), V_i^b(k), V_i^s\}_{i\in N}$, *population measures* $\{\hat{\mu}_i^o(k), \hat{\mu}_i^b(k), \mu_i^s\}_{i\in N}$, *the distribution of bonds across dealers* $\{s_i\}_{i\in N}$, *prices* $\left\{\hat{p}_{i,j}^{bid}(k), \hat{p}_{i,j}^{ask}(k), \hat{P}_{i,j}(k)\right\}_{i,j\in N}$, *and entry decisions* $\{\nu_i(k)\}_{i\in N}$ *such that*

1. *Value functions solve investors' optimization problems* (A.31)–(A.33).
2. *Population measures and the distribution of bonds across dealers solve inflow-outflow equations* (9), (A.30), *market clearing conditions* (6)–(7), *and interdealer transactions equations* (8).
3. *Prices arise from bargaining* (A.34)–(A.36).
4. *Entry decisions solve* (1) *and* $\sum_{i \in N} \nu_i(k) = 1$.

2 Main Results

Symmetric Equilibrium

The following proposition shows that a continuum of symmetric equilibria exists. We define an equilibrium as symmetric if dealers have an identical measure of buyers and sellers (even if the composition differs). A trivial example is when buyers choose all three dealers with the same probability: $\nu_i(k) = \frac{1}{3}$ for all k. That is, dealers—instead of specializing—serve the entire spectrum of customers from buy-and-hold to liquidity investors. Importantly, in the symmetric equilibria, dealers have identical network centrality. This shows that, first, we do not have any baked-in dealer heterogeneity. Second, client heterogeneity *alone* does not guarantee dealer heterogeneity.

Proposition 1 (Symmetric Equilibrium)**.** *A continuum of symmetric equilibria exists, where dealers have identical client masses:* $\mu_1^s = \mu_2^s = \mu_3^s$.

Asymmetric Specialization Equilibrium

We focus on the asymmetric equilibrium of Proposition 2. Without loss of generality, we label the dealer that endogenously attracts the clients with the slowest, intermediate, and greatest liquidity needs as dealer 1, 2, and 3, respectively.[12] Figure 4 illustrates the result.

Assumption 1. *Suppose*

$$\lambda_{\mathrm{DD}} z_{\mathrm{DD}} > \lambda_{\mathrm{D}} z_{\mathrm{D}}. \tag{11}$$

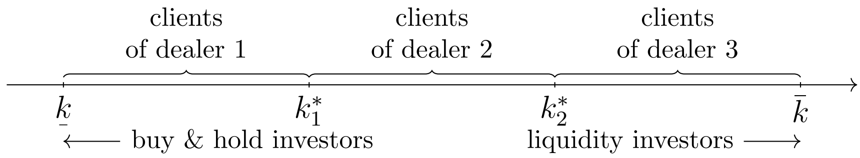
Lemma 1. *Suppose* k^* *is such that* $\hat{V}_i^b(k^*) = \hat{V}_j^b(k^*)$. *Then,* $\hat{V}_i^b(k) - \hat{V}_j^b(k)$ *is the same sign as* $(k - k^*)\left(\lambda_{\mathrm{DD}} z_{\mathrm{DD}} - \lambda_{\mathrm{D}} z_{\mathrm{D}}\right)\left(\mu_i^s - \mu_j^s\right)$.

Lemma 1 shows that if $\mu_i^s - \mu_j^s > 0$, buyers with $k > k^*$ prefer dealer i over dealer j. That is, relatively liquidity investors prefer the dealer with a larger seller client mass.

[12] Other asymmetric equilibria have identical properties, but with dealer indices reversed.

Proposition 2 (Asymmetric Specialization Equilibrium)**.** *An asymmetric equilibrium exists characterized by cutoffs* $\{k_1^*, k_2^*\}$*, where* $\underline{k} < k_1^* < k_2^* < \overline{k}$*, buyers with* $k < k_1^*$ *choose dealer 1, buyers with* $k \in [k_1^*, k_2^*]$ *choose dealer 2, and buyers with* $k > k_2^*$ *choose dealer 3. Buyers at the cutoff* $k = k_1^*$ *are indifferent between dealers 1 and 2:* $V_1^b(k_1^*) = V_2^b(k_1^*)$*, and buyers at the cutoff* $k = k_2^*$ *are indifferent between dealers 2 and 3:* $V_2^b(k_2^*) = V_3^b(k_2^*)$*.*[13]

Figure 4: Endogenous Cutoffs $\{k_1^*, k_2^*\}$ and Specialization in Customers



To explain how clients sort across dealers, we first characterize properties of the asymmetric equilibrium.

Proposition 3 (Properties of the Specialization Equilibrium)**.** *Suppose dealers* i *and* j *specialize in liquidity and buy-and-hold investors, respectively:* $i > j$*. Dealer* i *has a larger mass of buyers and sellers:* $\mu_i^b > \mu_j^b$ *and* $\mu_i^s > \mu_j^s$ *but fewer owners and bonds in circulation:* $\mu_i^o < \mu_j^o$ and $s_i < s_j$*. Dealer* i *demands a higher expected ask-price:* $\overline{p}_i^{ask}(k) > \overline{p}_j^{ask}(k)$ *for all* $k \in [\underline{k}, \overline{k}]$*. For the higher ask-price it charges, dealer* i*, in turn, offers a better return service:* $V_i^s > V_j^s$ *(by either buying back at a higher bid-price on average:* $\overline{p}_i^{bid} > \overline{p}_j^{bid}$*, executing sell orders more quickly* $m_i^b > m_j^b$*, or both).*

Here is how customers with different liquidity needs endogenously sort across different dealers. Buyers tradeoff a dealer's ask-price, $\overline{p}_i^{ask}(k)$, versus its return service, V_i^s.[14] Some dealers charge a high ask-price but, in return, offer a better service if the client has to return the bond: Depending on the parameter values, the dealer either buys back at a higher bid-price, executes the order more quickly, or both. Others charge a cheaper ask-price but offer a worse return service. Buyers who expect to reverse their position quickly (i.e., high k buyers) care more about what happens to them as a seller. They, as a result, choose the dealer based on its return service and are willing to

[13]Showing that the equilibrium is unique is tedius. We thus omit the uniqueness proof for now.

[14]The value function of dealer i's seller client, V_i^s, summarizes the dealer's return service. In the Appendix, we show that $V_i^s = \frac{r}{(r+m_i^b)}\left(\frac{\delta - x}{r}\right) + \frac{m_i^b}{(r+m_i^b)}\left(\overline{p}_i^{bid}\right)$. Thus, the expected utility of a seller, who is a client of dealer i, increases in both the probability of getting matched with a buyer and the expected bid-price.

pay the higher ask-price. Buy-and-hold investors, less concerned with turning into a seller later on, instead, choose the dealer offering the cheapest price. Figure 7 illustrates the tradeoff. Appendix B explains the tradeoff in detail.

The ask-prices dealers charge, as a result, serve as a sorting device. Dealers quoting a higher ask-price specialize in buyers who turn around and sell quickly; dealers quoting a cheaper price specialize in buy-and-hold investors. The difference in clientele, in turn, supports the heterogeneity in the value function of sellers, prices, and client masses across dealers.

The intuition for why a dealer specializing in liquidity investors offers a better value to its sellers is as follows. Substituting in the bid-prices, the value function of a seller is a weighted sum of the expected trading surpluses from "in-house" and "inter-house" matches:

$$rV_i^s = \delta - x + (\lambda_D z_D) \underbrace{\mu_i^b E_i^b\left[\omega_{ii}(k)\right]}_{\substack{\text{gains from} \\ \text{in-house matches}}} + (\lambda_{\text{DD}} z_{\text{DD}}) \underbrace{\sum_{j \in N_i} \mu_j^b E_j^b\left[\omega_{ji}(k)\right]}_{\substack{\text{gains from inter-} \\ \text{house matches}}}.$$

Due to assumption (11), the weight on the gains from inter-house matches is larger than the weight on the gains from in-house matches. The expected utility of a seller, as a result, depends more on the inter-house matches. For clients of a dealer specializing in liquidity investors, the inter-house matches are with buy-and-hold buyers. A match with a buy-and-hold investor, in turn, yields a larger trading surplus than a match with a liquidity investor because buy-and-hold investors are the natural investors in the bond.[15] Put together, a dealer specializing in liquidity investors offers a better value to its sellers. The better value manifests as either a higher bid-price, faster execution speed, or both. The mechanism reverses for dealers specializing in buy-and-hold investors.

An Endogenous Core-Periphery Network

We measure a dealer's network centrality by its volume of interdealer trades, M_i^{DD}, given in (5). Since, the number of links is identical across dealers, our measure is equivalent to: the number of links weighted by the strength of the link (that is, by the volume of trade between dealers). We thus define dealer i as more central (i.e., core) than dealer j if dealer i intermediates a larger volume of interdealer trades (M_i^{DD}) than dealer j.

Definition 1. *Dealers i and j are defined as relatively core versus peripheral*

[15]That is, buy-and-hold investors have a higher reservation value for the bond than liquidity investors: $V^o(k) - V^b(k)$ is decreasing in k.

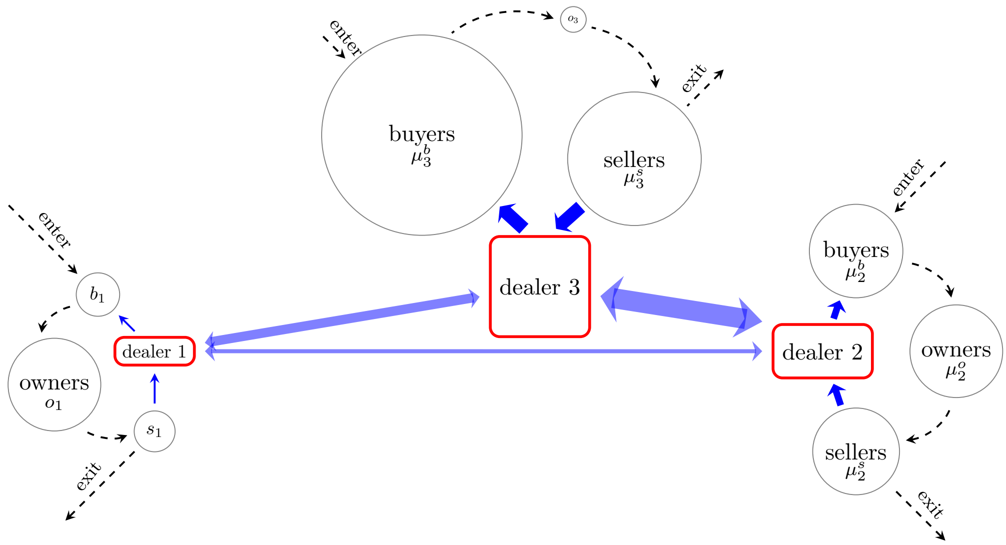
if $M_i^{DD} > M_j^{DD}$.

Proposition 4 gives the main insight of our paper: The heterogeneity in client masses across dealers translates to a heterogeneity in the network centrality across dealers. Dealers of liquidity investors—supported by their large client mass—intermediate larger volumes of interdealer trades and, consequently, form the core. The large client base of core dealers itself endogenously arises from the characteristics of clients that self-select with core dealers (namely, liquidity investors). The mechanism reverses for peripheral dealers. Figure 5 illustrates the result.

Proposition 4 (An Endogenous Core-Periphery Network)**.** *The dealers that attract more liquidity investors intermediate more CDC chains,* $M_i^D > M_j^D$*. They also intermediate more interdealer (i.e. CDDC) trades,* $M_i^{DD} > M_j^{DD}$*, and thus form the core.*

Figure 5: An Endogenous Core-Periphery Structure

The figure illustrates the equilibrium network structure in the asymmetric equilibrium. The equilibrium exhibits a core-periphery network. Dashed (black) lines represent clients' life-cycle between different types (buyer, owner, and seller). Solid (blue) lines represent bond transaction flows. The sizes of circles represent the sizes of client measures.



Key Ingredients

The endogenous dealer heterogeneity relies on three ingredients. The first ingredient is matching frictions ($\lambda < \infty$) together with an imperfectly competitive dealer market. Absent trading frictions ($\lambda \to \infty$), the dealer heterogeneity and, hence, the core-periphery structure do not arise.

The second ingredient is the parameter condition in (11): $\lambda_{\text{DD}} z_{\text{DD}} > \lambda_{\text{D}} z_{\text{D}}$. It says that, for a dealer heterogeneity to emerge, clients have to somehow benefit from interdealer intermediation chains and, consequently, prefer a dealer who relies relatively more on intermediation chains. Otherwise, they would either all pool with one dealer (consequently, only a monopoly dealer exists) or choose all dealers with the same probability (that is, only the symmetric equilibrium exists). The two ways to satisfy the condition are $\lambda_{\text{DD}} > \lambda_{\text{D}}$ and $z_{\text{DD}} > z_{\text{D}}$. The first says that two dealers are collectively more efficient in producing matches than if each worked on their own. The second says that clients extract a larger fraction of the trading surplus in two-dealer chains than in one-dealer chains. We abstract from potential microfoundations for why intermediation chains are beneficial. We, instead, capture them in a reduced form through (11). Glode and Opp (2014), for example, show that, in a model with adverse selection, when multiple dealers are involved in a chain, more trades take place than without intermediation chains. Their model, as a result, implies: $\lambda_{\text{DD}} > \lambda_{\text{D}}$. The main insight of our paper—that heterogenous clients endogenously sort across different dealers, and that specialization, in turn, supports dealer heterogeneity—does not depend on the underlying microfoundations that generate the parameter conditions.[16]

The third ingredient is dealer segmentation: a client can only sell through the dealer she initially chooses. If clients can later sell through any dealer, specialization would not arise. The dealer segmentation captures a fixed cost of building a client-dealer relationship that the client, then, needs to recoup over multiple subsequent trades. Presumably, such costs exist due to agency and contractual frictions, in the absence of which, clients would freely choose new dealers. Thus, our results suggest that the core-periphery phenomenon inherently arises from contractual frictions between OTC counterparties.[17]

The extent of all three ingredients increases the extent of dealer heterogeneity and, hence, the core-periphery structure. For example, as matching frictions increase, the extent of dealer heterogeneity and the core-periphery structure also increases.

3 Empirical Predictions

We now tie the network centrality results with the previous results on specialization. We highlight testable predictions of our model and, where available,

[16]Other model implications, however, depend on whether $\lambda_{\text{DD}} > \lambda_{\text{D}}$ or $z_{\text{DD}} > z_{\text{D}}$.

[17]The fact that the client segmentation is asymmetric—a buyer can choose over dealers, but a seller cannot—is immaterial.

compare them with the empirical evidence.

3.1 Client Trades

A broader interpretation of our model is: Core and peripheral dealers specialize in investment positions with short and long holding periods, respectively. For this interpretation, it does not matter if orders come from different clients or if the same client sends orders she expects to reverse quickly to a core dealer and her buy-and-hold positions to a peripheral dealer.

If we assume each order is tied to a different client, a narrower interpretation emerges: Peripheral and core dealers specialize in buy-and-hold and liquidity investors, respectively. In the paper, we focus on this interpretation. Liquidity investors could be, for example, investment funds that track indices and, hence, trade frequently, while buy-and-hold investors could be pension funds. A direct evidence for this prediction so far does not exist because in a typical dataset (such as that of LS and NHS) client identities are anonymous.[18]

CDC and CDDC chains A core dealer intermediates more CDC chains than a peripheral dealer:

$$M_c^D > M_p^D \tag{12}$$

Thus, core dealers account for a larger fraction of not only interdealer trades (hence their labels) but also client trades.[19] This result is not trivial. The core-periphery phenomenon is a statement about how dealers trade amongst each other, not how much they trade with clients. The phenomenon by itself, as a result, does not preclude other theories predicting that, for example, core dealers trade mainly with other dealers, and that peripheral dealers account for most of the client trades. Such theories would still be able to argue that they explain the core-periphery phenomenon. LS and NHS, however, document that core dealers also account for a larger fraction of client trades.[20] Thus, a convincing theory has to explain why core dealers account for a larger fraction of both interdealer and client trades. We not only reconcile the two facts but also show that core dealers' large volumes of client trades

[18]LS find that core dealers specialize in medium-size trades. The medium size trades, in turn, tend to flow from municipal mutual fund clients, who trade frequently. This finding is consistent with our mechanism.

[19]The total client trades of a dealer are $2M_i^D + M_i^{DD}$ (2 in front of M_i^D captures the fact that a CDC chain involves two client trades: the CD leg and DC leg). Thus, $M_c^D > M_p^D$ and $M_c^{DD} > M_p^{DD}$ imply that the total volume of client trades are larger for a core dealer.

[20]LS document that the top 5.4% dealers (by centrality) account for 75% of all client transactions.

are precisely why they form the core.

A core dealer intermediates more CDC chains both in levels as in (12) and as a fraction of all chains it intermediates:

$$\frac{M_c^D}{M_c^D + M_c^{DD}} > \frac{M_p^D}{M_p^D + M_p^{DD}} \tag{13}$$

Thus, a core dealer intermediates client trades more on its own than by relying on the interdealer market. A peripheral dealer, in contrast, relies more on other dealers and hence on long intermediation chains for its liquidity service to clients.[21] For a peripheral dealer, CDDC chains comprise a larger fraction of all its intermediations than for a core dealer:

$$\frac{M_p^{DD}}{M_p^D + M_p^{DD}} > \frac{M_c^{DD}}{M_c^D + M_c^{DD}}. \tag{14}$$

Eq. (13) also implies that the average chain length is longer for a peripheral dealer:

$$\frac{M_p^D}{M_p^D + M_p^{DD}}(1) + \frac{M_p^{DD}}{M_p^D + M_p^{DD}}(2) > \frac{M_c^D}{M_c^D + M_c^{DD}}(1) + \frac{M_c^{DD}}{M_c^D + M_c^{DD}}(2), \tag{15}$$

where inside the brackets are the chain lengths. LS and NHS document the same patterns as (12)-(15).

Execution speed The rate at which dealer i fills clients' buy orders is:

$$m_i^s \equiv \frac{M_i^D + 0.5 M_i^{DD}}{\mu_i^b} = \sum_{j \in N} \lambda_{ij} \mu_j^s. \tag{16}$$

The denominator is the total buy orders the dealer receives; the numerator is, out of the total, how many it executes. The ratio captures the fraction of all buy orders the dealer executes.[22] The rate of filling sell orders is analogously defined as:

$$m_i^b \equiv \frac{M_i^D + 0.5 M_i^{DD}}{\mu_i^s} = \sum_{j \in N} \lambda_{ij} \mu_j^b \tag{17}$$

Consider the difference between execution speeds of any two dealers i and

[21]To see this, multiply (12) by negative 1 and add 1 to both sides.

[22]I refer to the ratio as the probability of trade although they are intensities, not probabilities. More precisely, in a small time interval $[t, t + dt]$, the order is executed with probability $m_i^s dt$.

j:

$$m_i^\tau - m_j^\tau =$$

$$= \left[\lambda_D \mu_i^\tau + \lambda_{DD} \mu_j^\tau + \lambda_{DD} \sum_{j \in N/\{i,j\}} \mu_j^\tau\right] - \left[\lambda_D \mu_j^\tau + \lambda_{DD} \mu_i^\tau + \lambda_{DD} \sum_{j \in N/\{i,j\}} \mu_j^\tau\right]$$

$$= -(\lambda_{DD} - \lambda_D)\left(\mu_i^\tau - \mu_j^\tau\right)$$

for $\tau = \{s, b\}$. Thus, if $\lambda_{DD} < \lambda_D$, a core dealer executes at a faster rate: $m_c^\tau > m_p^\tau$ for $\tau = \{s, b\}$. If $\lambda_{DD} = \lambda_D$, a core dealer offers the same execution speed as a peripheral dealer: $m_i^\tau = m_j^\tau$. If $\lambda_{DD} > \lambda_D$, a core dealer executes at a slower rate: $m_c^\tau < m_p^\tau$. The intuition for the latter is: A core dealer fills large volumes of client orders (the numerator in (16) and (17)), but the volume of orders submitted to the dealer is even greater (the denominator). Peripheral dealers, in contrast, transact fewer client volumes, but the amount of orders they receive is even fewer.[23]

Transacted and quoted ask-price A core dealer transacts, on average, at a lower ask-price than a peripheral dealer:

$$E_c^b\left[\overline{p}_c^{ask}(k)\right] < E_p^b\left[\overline{p}_p^{ask}(k)\right] \tag{18}$$

Eq. (18) compares prices across dealers averaged in two dimensions. The first dimension, as discussed earlier, is across possible end-sellers a k-type buyer could be matched with: $\overline{p}_i^{ask}(k)$. The empirical counterpart to $\overline{p}_i^{ask}(k)$ would be dealers' effective quoted prices. Recall that, for a given k-buyer, a core dealer quotes a higher expected ask-price:

$$\overline{p}_c^{ask}(k) > \overline{p}_p^{ask}(k) \tag{19}$$

Eq. (19) is the counterfactual we observe in the model. The second dimension, captured by $E_i^b[]$, is across dealer i's equilibrium buyer mass. Because ask-prices, $\overline{p}_i^{ask}(k)$, decrease with the buyer type k, and a core dealer's clients are in equilibrium high k buyers, in a transaction price data, we would observe (18), not (19). Thus, (18) is the testable prediction relevant to transaction price data, not (19).[24]

[23]In the data (e.g. in LS and NHS), a dealer's execution speed is unobservable because—although its transaction volume (the numerator) is observable—the volume of orders it receives (the denominator) is not. Thus, a direct empirical evidence on dealers' execution speed is unavailable.

[24]In a data with both quoted and transaction prices, it is possible to check (19) also.

Transacted and quoted bid-price Whether a core dealer buys back at a higher bid-price depends on λ_D vs λ_{DD}. If λ_{DD} is sufficiently low, a core dealer buys back at a lower return (i.e. bid-) price than a peripheral dealer. The intuition is as follows. For a sufficiently low λ_{DD}, a core dealer is so fast that even if it buys back at a lower price, it still offers a greater overall value to its sellers than a peripheral dealer. Conversely, for a high value of λ_{DD}, a core dealer is slower. To compensate for its inferior speed, it has to offer a narrower bid-ask spread. It does so by buying back at a higher price than a peripheral dealer.

A dealer's quoted and transacted bid-prices coincide because sellers do not differ by their liquidity type and, hence, face the same bid-price.

Transacted bid-ask spread A core dealer charges, on average, a narrower bid-ask spread than a peripheral dealer:

$$E_c^b\left[\overline{\phi_c}(k)\right] < E_p^b\left[\overline{\phi_p}(k)\right], \tag{20}$$

where, similar to the discussion of average ask-prices, $E_i^b\left[\overline{\phi_i}(k)\right]$ is an average across two dimensions. Eq. (20) can be seen in Figure 8. The trading surplus (for the entire intermediation chain) decreases with the end-buyer's liquidity type k. Bid-ask spreads, as a result, also decrease with k because bid-ask spreads are proportional to the total gains from trade. This together with the fact a core dealer's buyers are high k buyers imply (20).

NHS document the same for the asset-backed securities market, but LS find the opposite with the municipal bond market data. Both studies also document that longer intermediation chains have wider bid-ask spreads. Consistent with this finding, our model predicts that the average chain involving a peripheral dealer is longer and that peripheral dealers charge clients wide spreads.

Quoted bid-ask spread Whether a core dealer also quotes a narrower bid-ask spread depends on if the core dealer executes orders at a faster rate and by how much faster. The latter, in turn, depends on λ_{DD}. Figure 8 shows how dealers' bid-ask spreads differ. For a high value of λ_{DD}, a core dealer is slower. To compensate for its inferior execution speed and to preserve $V_c^s > V_p^s$, a core dealer has to offer to buy back a higher bid-price than a peripheral dealer. For a sufficienty high λ_{DD}, the bid-price is so high that the core dealer offers a narrower bid-ask spread for all k.[25] On the other extreme,

[25]That is, the value of being a client of a core dealer comes from a narrower bid-ask spread, not execution speed.

for a relatively low λ_{DD}, a core dealer is faster. As a result, the core dealer's bid-price can be lower, and hence its quoted bid-ask spread wider for all k. For an intermediate values of λ_{DD}, at some $\tilde{k} \in [\underline{k}, \overline{k}]$, the two bid-ask spread curves cross so that $\overline{\phi_c}(k) > \overline{\phi_p}(k)$ for $k \in [\underline{k}, \tilde{k})$, and $\overline{\phi_c}(k) < \overline{\phi_p}(k)$ for $k \in (\tilde{k}, \overline{k}]$.[26]

3.2 Interdealer Trades

We now discuss the roles that core and peripheral dealers play on the interdealer market. Below results are novel testable predictions. Appendix A characterizes prices $\{P^{bid}, P^{ask}\}$, bid-ask spreads Φ, and execution speed dealers' face from each other.

Proposition 5 (Prices and Liquidity Provision on the Interdealer Market)**.** *Suppose dealers indexed c and p are relatively core and peripheral dealers, respectively. A core dealer charges other dealers a higher ask-price, $P_c^{ask} > P_p^{ask}$, buys back at a lower bid-price, $P_c^{bid} < P_p^{bid}$, and hence charges other dealers a wider bid-ask spread, $\Phi_c > \Phi_p$, than a peripheral dealer. A core dealer buys and sells more than a peripheral dealer: $\lambda_{\text{DD}}\mu_d^s\mu_c^b > \lambda_{\text{DD}}\mu_d^s\mu_p^b$ and $\lambda_{\text{DD}}\mu_d^b\mu_c^s > \lambda_{\text{DD}}\mu_d^b\mu_p^s$. A core dealer provides a faster execution speed: $\lambda_{\text{DD}}\mu_c^\tau > \lambda_{\text{DD}}\mu_p^\tau$ for $\tau = \{s, b\}$.*

Core dealers—supported by the large volumes of clients' orders—supply liquidity to other dealers. They do so in two ways. First, they transact greater volumes.[27] The number of bonds an arbitrary dealer d sells to another dealer i is $\lambda_{\text{DD}}\mu_d^s\mu_i^b$, and the number of bonds it buys from dealer i is $\lambda_{\text{DD}}\mu_i^s\mu_d^b$. Since a core dealer has a larger client mass, dealer d trades proportionally more with a core dealer on both sides of the trade. Second, a core dealer offers a faster execution speed to other dealers. The rate at which dealer i fills dealer d's sell orders is

$$\frac{\lambda_{\text{DD}}\mu_d^s\mu_i^b}{\mu_d^s} = \lambda_{\text{DD}}\mu_i^b.$$

Thus, the execution speed of dealer i is proportional to its client size. Since a core dealer has a larger buyer mass, it executes dealer d's orders more quickly:

$$\lambda_{\text{DD}}\mu_c^b > \lambda_{\text{DD}}\mu_p^b.$$

[26]In other words, as λ_{DD} decreases, the cutoff at which the two bid-ask spreads cross increases.

[27]This holds by construction because we define a dealer's network centrality by its total interdealer volume.

It is analogous for dealer d's buy-side trades.[28]

For the liquidity they provide, core dealers charge other dealers wide bid-ask spreads, $\Phi_c > \Phi_p$, due to two effects. First, when a dealer buys from a core dealer, the dealer ultimately buys from an end-seller who has a high reservation value.[29] The end-seller's high reservation value, in turn, manifests as a high interdealer ask-price.[30] Thus, from the perspective of a dealer, it is more expensive to buy from a core dealer than from a peripheral dealer: $P_c^{ask} > P_p^{ask}$. Second, on the reverse trip, when a dealer sells back to a core dealer, the dealer ultimately sells to liquidity investors (high k buyers), who have low reservation values. A dealer, as a result, sells back at a lower (bid-) price to a core dealer, $P_c^{bid} < P_p^{bid}$. Put together, a dealer faces a wider bid-ask spread from a core dealer. Recall that the opposite holds for client transactions: Core dealers charge clients narrower bid-ask spreads (on average, across its buyers).[31]

Bonds, as a result, cycle through the economy starting with, say, a core dealer's client, then the interdealer network, and eventually end with buy-and-hold investors who are concentrated with peripheral dealers. The cycle repeats when a buy-and-hold investor gets a liquidity shock and sells the bond. The sell order primarily gets absorbed, via the interdealer network, first by core dealers and their clients. Thus, core dealers serve as a central conduit in transmitting assets through the economy from one end-customer to another. Peripheral dealers consume the liquidity core dealers supply and pass it down to their clients.

[28] As in the earlier discussion of liquidity immediacy from clients' perspective, because the amount of orders dealers receive is unobservable (whether from clients or other dealers), we lack a direct empirical evidence on liquidity immediacy.

[29] Recall that sellers of a core dealer have a higher value function $V_c^s > V_p^s$ and, hence, a higher reservation value for the bond.

[30] The interdealer price between any two dealers is the average between the end-buyer and end-seller reservation values.

[31] LS consider how dealers split the total round-trip spread between prices at the CD to DC legs and find that dealers closer to the end-buyer extract a bigger fraction of the total spread. They, however, do not focus on how core vs. peripheral dealers split the intermediation surplus. NHS consider similar splits and conclude that core dealers take a narrower chunk of the total spread. In contrast, we characterize bid-ask spreads from dealers' perspective to understand the liquidity service core vs. peripheral dealers provide other dealers.

4 Additional Results

4.1 Dealer Interconnectedness

In this section, we contrast environments with and without the interdealer market and show that dealer interconnectedness increases customers' welfare, dealer profits, bond liquidity, and bond prices. Without the interdealer market, dealers intermediate between only their own customers. We assume the supply of bonds circulating among customers of each dealer is identical at $s_i = S/3$. The environment without the interdealer market is similar to Vayanos and Wang (2007).[32] Markets in their setting are the counterparts to dealers in our setting.

How clients sort in the absence of interdealer trades is identical to the environment with interdealer trades. Buyers tradeoff a dealer's ask price versus its return service. Buy-and-hold investors choose the dealer offering a cheaper price, while liquidity investors choose the dealer offering a better return service. The dealer specializing in liquidity investors has a larger buyer and seller client mass.

Interconnectedness has two effects. First, the sorting mechanism is more general. Without interdealer trades, the larger dealer offers a faster execution speed and buys back at a higher bid-price. With interdealer trades, in contrast, whether the larger (core) dealer is faster and, consequently, whether its return price is higher depends on parameter values. Second, interconnectedness decreases the dispersion in prices and reservation values across dealers. With or without interdealer trades: $\overline{p}_2^{ask}(k) > \overline{p}_1^{ask}(k)$, where $i = 2$ is the larger dealer. Without interdealer trades, the difference, however, is large enough that the average price across buyers is higher—$E_2^b\left[\overline{p}_2^{ask}(k)\right] > E_1^b\left[\overline{p}_1^{ask}(k)\right]$—eventhough the larger dealer specializes in high k-buyers and prices decrease with k. With interdealer trades, the difference in prices across dealers (that is, for the same buyer) decreases and, as a result, $E_2^b\left[\overline{p}_2^{ask}(k)\right] < E_1^b\left[\overline{p}_1^{ask}(k)\right]$.[33]

We define customers' welfare as

$$W^C \equiv \sum_{i\in N}[\int_{\underline{k}}^{\overline{k}} \hat{\mu}_i^b(k)V_i^b(k)dk + \int_{\underline{k}}^{\overline{k}} \hat{\mu}_i^o(k)V_i^o(k)dk + \mu_i^s V_i^s. \qquad (21)$$
$$+ \frac{1}{r}\int_{\underline{k}}^{\overline{k}} V_i^b(k)\hat{f}(k)\nu_i(k)dk]$$

[32]Note that Vayanos and Wang (2007) is a special case with $z_{ij} = 1$ and $N_i = \{\emptyset\}$ for all i.

[33]Interconnectedness has a similar effect on buyer's reservation values.

For dealer i, the present value of the stream of flow profits is

$$
\begin{aligned}
W_i^D \equiv & \frac{1}{r}\int_{\underline{k}}^{\overline{k}} \lambda_{\mathrm{D}}\hat{\mu}_i^b(k)\mu_i^s\,(1-2z_{\mathrm{D}})\left(V_i^o(k)-V_i^b(k)-V_i^s\right)dk \qquad (22)\\
&+\frac{1}{r}\sum_{j\in N_i}\left(\int_{\underline{k}}^{\overline{k}} \lambda_{\mathrm{DD}}\hat{\mu}_i^b(k)\mu_j^s\left(\frac{1-2z_{\mathrm{DD}}}{2}\right)\left(V_i^o(k)-V_i^b(k)-V_j^s\right)dk\right)\\
&+\frac{1}{r}\sum_{j\in N_i}\left(\int_{\underline{k}}^{\overline{k}} \lambda_{\mathrm{DD}}\hat{\mu}_j^b(k)\mu_i^s\left(\frac{1-2z_{\mathrm{DD}}}{2}\right)\left(V_j^o(k)-V_j^b(k)-V_i^s\right)dk\right).
\end{aligned}
$$

The first term captures profits from intermediations directly between its customers (that is, CDC chains). The second and third terms are profits from buy and sell interdealer transactions, respectively (that is, CDDC chains). The total profit across dealers is

$$W^D \equiv \sum_{i\in N} W_i^D. \qquad (23)$$

The total welfare of all agents in the economy is then

$$W_{all} \equiv W^C + W^D. \qquad (24)$$

As Proposition 6 shows, the total welfare depends only on the aggregate mass of sellers:

$$\mu_N^\tau \equiv \sum_{i\in N}\mu_i^\tau \qquad \text{for } \tau=\{s,b\}.$$

Proposition 6. *The total welfare is given by*

$$W_{all} = \frac{\delta}{r}S - \frac{x}{r}\mu_N^s. \qquad (25)$$

The first term is the present value of the stream of bond coupon flows. The welfare in a frictionless environment corresponds to this term because only investors that enjoy the full value of the coupon flow own the bond. Matching frictions, however, create misallocations: investors (with total mass μ_N^s) who dislike holding the bond (recall the disutility, x) own the bond also. Thus, the second term represents the welfare loss from matching frictions.

Proposition 7 (The Effect of Interconnectedness)**.** *Customers' welfare (W^C) and the total welfare (W_{all}) increase with dealer interconnectedness.*

The presence of the interdealer market improves bond liquidity: it increases the aggregate volume of transactions, narrows bid-ask spreads, and speeds up transaction times. Greater liquidity, in turn, alleviates misallocations: a larger number of investors who enjoy the full value of the coupon flow (hence, fewer sellers) own the bond. The more efficient asset allocation increases both the customer and total welfare.[34]

Second, since bonds are held proportionately more by investors with the greatest utility for them, bond prices increase and, in particular, approach the frictionless price. For the parameter values in Table 1, the measure of buyers is greater than the total bond supply; consequently, buyers are the marginal investors in the bond. In a frictionless environment ($\lambda \to \infty$), the bond price is the present value of buyers' valuation of the bond, $p = \frac{\delta}{r}$. With frictions, low-valuation investors also hold the bond, leading to discounted bond prices relative to the frictionless price. Thus, the more efficient allocation of bonds and the increase in bond prices imply that bond prices approach the frictionless price.

Fourth, if we proxy a dealer's inventory balance with its seller-to-buyer ratio, dealers achieve what looks like a full inventory risk-sharing. Without the interdealer market, the seller-to-buyer ratio differs across dealers and is higher for dealers that cater to buy-and-hold investors. With the interdealer market, as Proposition 8 shows, the ratio is identical across dealers. Lastly, interconnectedness decreases the dispersion of prices and liquidity across dealers.

Proposition 8. *In the presence of the interdealer market, the inventory balance is identical across dealers: for all $i \in N$,*

$$\frac{\mu_i^s}{\mu_i^b} = \frac{\mu_N^s}{\mu_N^b}. \tag{26}$$

4.2 Market Fragmentation

In this section, we analyze how interdealer market fragmentation affects customer welfare, dealer profits, and bond liquidity.[35] Keeping the level of interconnectedness fixed, we capture market fragmentation with the aggregate number of dealers in the economy, denoted by n_N. In particular, we compare three environments with an increasing aggregate number of dealers: (1) one dealer (that is, dealers are merged into one), (2) two dealers (dealers are merged into two), and (3) the benchmark environment with all three dealers. In the latter two cases, since multiple equilibria exist, we compare across only the asymmetric equilibrium of each environment. In the environment with

[34] Whether dealer profits increase depends on how the parameter condition (11) is satisfied.

[35] In this section, we assume that condition (11) is satisfied via $\lambda_{\text{DD}} > \lambda_{\text{D}}$.

just one dealer, the supply of bonds circulating among the dealer's clients is simply the aggregate supply of bonds, S.

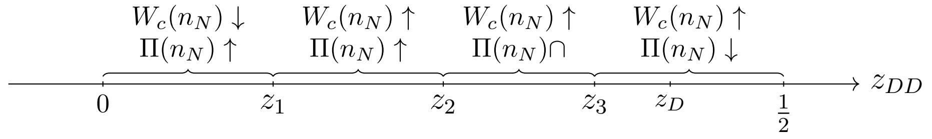
Proposition 9. *Increasing the aggregate number of dealers decreases the aggregate mass of sellers, μ_N^s, increases the aggregate volume of trade, $\sum_{i \in N} \left(M_i^D + M_i^{DD} \right)$, and increases the total welfare in the economy, W_{all}.*

Thus, market fragmentation alleviates misallocations in the economy. Increasing the aggregate number of dealers increases the length of an average intermediation chain in the economy. Since, by assumption, multiple dealers are more efficient in producing matches, aggregate transaction volumes increase. In turn, the efficiency of asset allocation and the total welfare increase.

Proposition 10 (The Effect of Market Fragmentation on the Welfare Split)**.** *Fixing clients' bargaining power in one-dealer intermediation chains, z_{D}, consider four regions of z_{DD} (clients' bargaining power in two-dealer chains): $0 < z_1 < z_2 < z_3 < \frac{1}{2}$. Customers' welfare decreases with market fragmentation (i.e. $W^C(n_N + 1) < W^C(n_N)$) in $z_{\text{DD}} \in (0, z_1]$ and increases in $z_{\text{DD}} \in (z_1, \frac{1}{2}]$. Dealers' profits increase with market fragmentation (i.e. $W^D(n_N + 1) > W^D(n_N)$) in $z_{\text{DD}} \in (0, z_2]$, non-monotone and concave in $z_{\text{DD}} \in (z_2, z_3]$, and decreases in $z_{\text{DD}} \in (z_3, \frac{1}{2}]$.*

How clients and dealers split the total welfare depends on whether clients' bargaining power increases or decreases with the chain length. Fixing the clients' bargaining power in one-dealer intermediation chains, z_{D}, consider four regions of z_{DD} (the clients' bargaining power in two-dealer chains), shown in Figure 6. Suppose, for example, $z_{\text{DD}} \geq z_{\text{D}}$ so that clients' bargaining power increases with the chain length. Then, Proposition 10 shows that, by lengthening the intermediation chain, clients collectively tilt the gains from trade in their favor at the expense of dealers. And the most fragmented interdealer market yields the largest customer welfare. Dealers instead prefer for other dealers to exit so that the interdealer market is as concentrated as possible. Conversely, if $z_{\text{DD}} < z_1 < z_{\text{D}}$ so that clients' bargaining power decreases with the chain length, dealer profits increase with market fragmentation but at the expense of customer welfare.

Figure 6: Regions of Clients' Bargaining Power in 2-dealer Chains, z_{DD}



Consider now the effect of fragmentation on bond prices and the bid-ask spreads that clients face. To compare prices across different environments with different network structures, we take the weighted average across dealers:

$$\bar{p}_{ask} \equiv \frac{1}{\sum\limits_{i\in N}\left(\frac{1}{2}M_i^{DD}+M_i^D\right)}\sum_{i\in N}\left[\left(\frac{1}{2}M_i^{DD}+M_i^D\right)p_i^{ask}\right] \tag{27}$$

$$\bar{p}_{bid} \equiv \frac{1}{\sum\limits_{i\in N}\left(\frac{1}{2}M_i^{DD}+M_i^D\right)}\sum_{i\in N}\left[\left(\frac{1}{2}M_i^{DD}+M_i^D\right)p_i^{bid}\right] \tag{28}$$

$$\bar{\phi} \equiv \frac{1}{\sum\limits_{i\in N}\left(\frac{1}{2}M_i^{DD}+M_i^D\right)}\sum_{i\in N}\left[\left(\frac{1}{2}M_i^{DD}+M_i^D\right)\phi_i\right]. \tag{29}$$

Bond prices increase with market fragmentation, reflecting the fact that bonds are allocated more efficiently and held by investors with the greatest utility for them. The effect on bid-ask spreads, however, similar to the effect on dealer profits and customer welfare, depends on whether clients' bargaining power increases or decreases with the chain length. In particular, the direction of the effect is the same as for dealer profits. For example, in the regions of z_{DD} where dealers profits decrease, the bid-ask spreads clients face decrease with market fragmentation (consequently, with the average chain length) and reaches the minimum in the environment with three dealers.

4.3 Welfare Analysis

In this section, we analyze the social welfare in the asymmetric and symmetric equilibria and contrast them with the socially optimal amount of dealer specialization.[36] For exposition, we do so for a two-dealer environment. We start by denoting the cutoff k^*_{sym} such that the two dealers are identical: $\mu_1^s = \mu_2^s$. Decreasing the cutoff below k^*_{sym} increases dealer heterogeneity: It increases the measure of buyers choosing dealer 2 and, consequently, dealer 2's masses of buyers and sellers. We denote the cutoff that maximizes the total welfare W_{all} by k^*_{wel} and the actual equilibrium cutoff by k^*_{asym}. The following results are illustrated in Figure 9.

Proposition 11. *Dealer specialization is socially optimal:* $k^*_{wel} < k^*_{sym}$.

Proposition 11 implies that a core-periphery network is socially desirable. Specifically, the socially optimal cutoff prescribes dealer heterogeneity. The intuition is as follows. Buy-and-hold investors are the most natural owners of the bond. The quicker they can buy a bond and turn into an owner,

[36] In this section, we assume that condition 11 is satisfied via $\lambda_{\text{DD}} > \lambda_{\text{D}}$.

the more efficient is the asset allocation in the economy. In the symmetric equilibrium, every buyer faces the same probability of buying, irrespective of her liquidity type k or her dealer choice (i.e. the probability of finding a seller is a flat function of k). A social planner can pareto improve on this by tilting the probability of finding a seller so that the buy-hold investors buy more quickly. Dealer specialization achieves precisely that. A dealer specializing in buy-and-hold investors provides a faster liquidity immediacy than a dealer specializing in liquidity investors.

Proposition 12. *Relative to the social optimum, the equilibrium dealer heterogeneity and specialization are excessive:* $k^*_{asym} < k^*_{wel}$.

Proposition 12 implies that, although a core-periphery structure is socially desirable, the extent of the equilibrium core-periphery structure is excessive. Specifically, in the asymmetric equilibrium, buyers concentrate too much with the core dealer. The intuition is as follows. Sellers' incentives are aligned with that of the social planner: they prefer the seller-to-buyer ratio in the economy to be as small as possible.[37] Buyers, however, prefer more sellers in the economy because a greater number of potential counterparties translates to a greater bargaining power. And it is buyers who choose over dealers. In particular, buyers do not fully internalize the effect of their dealer choice on sellers because they receive only a fraction of the total gains from trade. If buyers were to extract a larger fraction of the intermediation surplus, their incentives align more closely with that of the social planner. Thus, both the asymmetric and the symmetric equilibria are inferior to the first best allocation: in the asymmetric equilibrium, buyers concentrate too much with one dealer (dealer heterogeneity is excessive), while, in the symmetric equilibrium, buyers concentrate too little (dealer heterogeneity is too little).

The natural next step is comparing the welfare of the asymmetric and symmetric equilibria. The next proposition shows that if buyers extract a sufficiently large fraction of the total gains from trade, then the welfare in the asymmetric equilibrium is higher than in the symmetric equilibrium.[38] This is because if they extract a larger fraction, they collect a larger fraction of any increase in the total welfare. Their incentives on dealer choice, as a result, align more closely with the social planner's. Thus, for a sufficiently large buyer bargaining power, the equilibrium featuring a core-periphery network dominates the equilibrium exhibiting a circular network.

[37] Recall that maximizing the social welfare is equivalent to minimizing the aggregate measure of sellers, which captures misallocations in the economy.

[38] The threshold $\bar{z}$ is such that $W_{all}(k^*_{asym}) = W_{all}(k^*_{sym})$.

Proposition 13. *The asymmetric equilibrium pareto dominates the symmetric equilibrium if buyers have a sufficiently large bargaining power:* $W_{all}(k^*_{asym}) > W_{all}(k^*_{sym})$ *if* $z_{ij} > \bar{z}$.

5 Assumptions

In this section, we discuss our assumptions and how relaxing them would affect our results. In Section 2, we discussed the key assumptions that our main results rely on. Relaxing below assumptions would make the environment more realistic but would not affect our main insights.

We assume a fully connected dealer network and that dealers do not choose who to connect to. Implicitly, we assume a zero cost of both initially connecting and maintaining the connection. We could relax this by assuming that dealers pay for an access to other dealers' clients. If dealers charge a cost per client, then we expect our results to remain the same. But if dealers charge a fixed amount regardless of the client size, dealers would pay only for an access to core dealers' clients. Our basic mechanism would go through, and the core-periphery structure would be even more pronounced. Although important, we leave for future work showing pairwise and group stability properties of the dealer networks in our model.

We take the aggregate number of dealers as fixed and do not model dealer entry and exit. We could model dealer entry as follows. Dealers have an outside opportunity. Dealers enter until the marginal dealer is indifferent between its outside opportunity and the profit it expects to make as one of the dealers in the economy. Nevertheless, endogenizing dealer entry would not change our main insight on dealer specialization.

We assume that dealers do not hold an inventory and that bonds sit on the balance sheet of client-sellers. We can recast the model so that, instead of clients holding the bonds on their balance sheet, dealers hold the bonds in their inventory. When a bond owner gets a liquidity shock and wants to sell her bond, she sells immediately to her dealer. The dealer, in turn, holds the bond in its inventory until it can match the bond with a buyer. With this interpretation, a dealer's inventory size would be proportional to its seller client size, and a core dealer, as a result, would have a larger inventory.

In our model, intermediation chains involve at most two dealers. Although we observe longer chains in the data, LS document that CDC and CDDC chains together comprise 90% of all intermediation chains and that the average intermediation chain involves just one dealer. Thus, our environment captures a majority of transactions. Nevertheless, we mention two ways to allow for longer intermediations. First, in our matching function specification, for a dealer to be involved in a chain, one of the end-customers has to be the dealer's own client. If, instead, a dealer can produce matches among clients of other dealers, intermediation chains can be longer than just two dealers. The second way is to allow dealers to hold inventory. In both ways, the longest chain in the model can be as long as the aggregate number of dealers in the model.

We assume a full information structure. In particular, dealers know client types, and clients know both their own and other dealers' client structure. The latter is reasonable since clients can figure out whether a dealer-brokerage firm is a large or small market player and, hence, a relatively core versus peripheral dealer. Regarding dealers' information on client types, Vayanos and Wang (2007) show that a clientele effect still emerges in the presence of asymmetric information about buyers' type. Thus, we predict that our main insight on dealer specialization would hold in the presence of asymmetric information.

We abstract from adverse selection problems. We observe the core-periphery structure and intermediation chains in markets where adverse selection problems are small. Currency and municipal bonds markets are an example. Thus, adverse selection problems cannot be a first order in explaining the core-periphery structure.

6 Conclusion

The network structure of over-the-counter markets exhibits a core-periphery structure: few dealers are highly interconnected with a large number of dealers, while a large of number of small dealers are sparsely connected. We build a search-based model of dealer network formation and show that the core-periphery structure emerges from dealer specialization. The dealers that attract a clientele of liquidity investors have a larger customer base, support a greater fraction of interdealer transactions, and, thus, form the core. The dealers that instead cater to buy-and-hold investors form the periphery.

A Client Masses, Value Functions, Prices

Client Masses The mass of k-type owners is given by

$$\left(\sum_{j\in N}\lambda_{ij}\mu_j^s\right)\hat{\mu}_i^b(k) = k\hat{\mu}_i^o(k). \tag{A.30}$$

The left-hand side is the flow of buyers that turn into a k-type owner of dealer i; the right-hand side reflects the flow of owners that experience a liquidity shock and switch to a seller.

Value Functions After simplifying and taking the continuous time limit of (10), we get

$$rV_i^b(k) = k\left(0 - V_i^b(k)\right) + \sum_{j\in N}\lambda_{ij}\mu_j^s\left(V_i^o(k) - V_i^b(k) - \hat{p}_{j,i}^{ask}(k)\right). \tag{A.31}$$

Inside the summation, if $j = i$, the transaction is with another customer of the same dealer. If $j \neq i$, the transaction instead involves an interdealer intermediation chain, and the end-seller is a customer of another dealer j. Analogously, the expected utility of a k-type bond owner who is a customer of dealer i is given by

$$rV_i^o(k) = \delta + k\left(V_i^s - V_i^o(k)\right). \tag{A.32}$$

The expected utility of a seller who is a customer of dealer i is given by

$$rV_i^s = \delta - x + \sum_{j\in N}\left(\int_{\underline{k}}^{\overline{k}}\lambda_{ij}\hat{\mu}_j^b(k)(\hat{p}_{i,j}^{bid}(k) - V_i^s)dk\right). \tag{A.33}$$

Characterization of Prices Specific to an Intermediation Chain

We first characterize prices specific to an intermediation chain (that is, specific to dealers and customers involved in a chain). We denote interdealer prices with capital letters (P) and client-to-dealer prices with small letters (p). A seller-client of dealer i sells to his dealer at the bid-price

$$\hat{p}_{i,j}^{bid}(k) = (1 - z_{ij})V_i^s + z_{ij}(V_j^o(k) - V_j^b(k)) \tag{A.34}$$

when the end-buyer is a k-type buyer of dealer j. Dealer i turns around and sells to dealer j at the interdealer price:

$$\hat{P}_{i,j}(k) = \frac{1}{2}V_i^s + \frac{1}{2}(V_j^o(k) - V_j^b(k)). \tag{A.35}$$

Dealer j, in turn, sells to its buyer-client at the ask price:

$$\hat{p}_{i,j}^{ask}(k) = z_{ij}V_i^s + (1 - z_{ij})\,(V_j^o(k) - V_j^b(k)). \tag{A.36}$$

If $j = i$, the end-buyer and seller are clients of the same dealer i, and the interdealer price $\hat{P}_{i,j}(k)$ is irrelevant. If $j \neq i$, the bond transaction instead involves an interdealer trade, and the end-buyer and seller are customers of different dealers.

Expected Prices and Liquidity, Clients' Perspective We now characterize the *expected* prices, expected bid-ask spreads, and probability of trade that clients face from their dealers. Averaging across all possible end-sellers that a buyer could be matched with, a k-type buyer-client of dealer i expects to buy at:

$$\overline{p}_i^{ask}(k) \equiv \frac{1}{m_i^s}\sum_{j\in N}\lambda_{ij}\mu_j^s\hat{p}_{j,i}^{ask}(k) \tag{A.37}$$

We define the liquidity immediacy buyers of dealer i face as:

$$m_i^s \equiv \sum_{j\in N}\lambda_{ij}\mu_j^s.$$

Analogously, the liquidity immediacy sellers of dealer i face is

$$m_i^b \equiv \sum_{j\in\{i,N_i\}}\left(\int_{\underline{k}}^{\overline{k}}\lambda_{ij}\hat{\mu}_j^b(k)\right)$$

Averaging across buyers of dealer i, an average buyer of dealer i expects to buy at:

$$p_i^{ask} \equiv E_i^b\left[\hat{p}_i^{ask}(k)\right], \tag{A.38}$$

where the expectation is over the buyer population measure.[39]

The price a seller of dealer i expects to sell at is the weighted average price across all buyers that she could be matched with (that is, buyers of

[39] In particular, for some function $f(k)$, $E_i^b\left[f(k)\right] \equiv \int_{\underline{k}}^{\overline{k}}\frac{\hat{\mu}_i^b(k)}{\mu_i^b}f(k)dk$.

both dealer i and dealer i's connections):

$$\overline{p}_i^{bid} \equiv \frac{1}{m_i^b} \sum_{j \in N} \lambda_{ij} \mu_j^b E_j^b [\hat{p}_{i,j}^{bid}(k)], \tag{A.39}$$

where $E_j^b[\hat{p}_{i,j}^{bid}(k)]$ is the weighted average price across buyers of dealer j.

We define the expected round-trip transaction cost from the perspective of a k-type buyer of dealer i as the expected ask price minus the expected bid price normalized by the mid-point:

$$\overline{\phi}_i(k) \equiv \frac{\overline{p}_i^{ask}(k) - \overline{p}_i^{bid}}{0.5(\overline{p}_i^{ask}(k) + \overline{p}_i^{bid})}. \tag{A.40}$$

Similarly, the round-trip transaction cost that an average buyer of dealer i expects is:

$$\phi_i \equiv \frac{p_i^{ask} - \overline{p}_i^{bid}}{0.5(p_i^{ask} + \overline{p}_i^{bid})}. \tag{A.41}$$

LS and NHS compute bid-ask spreads as follows. For a CDDC chain, for example, the bid-ask spreads clients face is the transaction price at the DC leg of the chain (i.e. the price a client buys at) minus the price at the CD leg (i.e. the price a client sells at) normalized by the mid-point in NHS and by the price at the CD leg in LS. LS regress this bid-ask spreads on the centrality of the first dealer.

Motivated by how clients in our model choose dealers, we instead take the perspective of a client of a particular dealer. We first take all chains j such that $\{j : CD_jD_iC\}$, i.e. chains where the buyer is a client of a dealer i, regardless of where dealer i finds the bond (other dealers, core vs peripheral, or its own clients). Averaging the price at the D_iC leg—across the chains in this set—gives the expected price a buyer of dealer i expects to buy at, again regardless of where the bond comes from. Second, we do the same on the CD leg: average the price at the CD_i leg across chains j such that $\{j : CD_iD_jC\}$. The average gives the expected selling price for a seller-client of dealer i. The bid-ask spread is the difference normalized by the midpoint. The difference in the computations matters only for chains longer than CDC and any averages computed using both short and long chains. Since CDC chains comprise a majority of all chains, our results are comparable to the results of LS and NHS.

Expected Prices and Liquidity, Dealers' Perspective We characterize now expected prices and bid-ask spreads that an arbitrary dealer, indexed d, faces from another dealer i. We denote prices and bid-ask spreads from interdealer transactions with capital letters, P and Φ, to contrast them from client-to-dealer transactions, p and ϕ.

Dealer d buys from dealer $i \in N_d$ at price $\hat{P}_{i,d}(k)$, defined in (A.35), if dealer d's client is a k-type buyer. The weighted average price across all buyers of dealer d is

$$P_i^{ask} = E_d^b[\hat{P}_{i,d}(k)]. \tag{A.42}$$

Conversely, dealer d sells to dealer i at price $\hat{P}_{d,i}(k)$ if dealer i's client is a k-type buyer. The weighted average price across buyers of dealer i is

$$P_i^{bid} = E_i^b[\hat{P}_{d,i}(k)]. \tag{A.43}$$

We define the bid-ask spread as the expected purchase price minus the expected selling price normalized by the midpoint:

$$\Phi_i = \frac{P_i^{ask} - P_i^{bid}}{0.5P_i^{ask} + 0.5P_i^{bid}}. \tag{A.44}$$

Although P_i^{ask}, P_i^{bid}, and Φ_i are specific to dealer d, for exposition, we suppress their dependence on d.

B Choosing Over Dealers, in Detail

We now explain in detail how investors sort across dealers for general λ_{DD} and λ_{D} without assuming their relative magnitudes. We first derive the expected utility of a buyer, owner, and seller client.

Consider the buyer's value function:

$$\begin{aligned}
rV_i^b(k) &= k\left(0 - V_i^b(k)\right) + \sum_{j\in\{i,N_i\}} \lambda_{ji}\mu_j^s \left(V_i^o(k) - V_i^b(k) - \hat{p}_{j,i}^{ask}(k)\right). \\
&= k\left(0 - V_i^b(k)\right) + \sum_{j\in\{i,N_i\}} \lambda_{ji}\mu_j^s[V_i^o(k) - V_i^b(k)] - \sum_{j\in\{i,N_i\}} \lambda_{ji}\mu_j^s\hat{p}_{j,i}^{ask}(k) \\
&= k\left(0 - V_i^b(k)\right) + [V_i^o(k) - V_i^b(k)]\left[\sum_{j\in\{i,N_i\}} \lambda_{ji}\mu_j^s\right] - m_i^s\frac{1}{m_i^s}\sum_{j\in\{i,N_i\}} \lambda_{ji}\mu_j^s\hat{p}_{j,i}^{ask}(k) \\
&= k\left(0 - V_i^b(k)\right) + [V_i^o(k) - V_i^b(k)]m_i^s - m_i^s\overline{p}_i^{ask}(k)
\end{aligned}$$

Then,

$$V_i^b(k) = \frac{V_i^o(k) m_i^s - m_i^s \overline{p}_i^{ask}(k)}{r + k + \overline{\mu}_i^s}$$
$$= \frac{r+k}{r+k+m_i^s}(0) + \frac{m_i^s}{r+k+m_i^s}\left(V_i^o(k) - \overline{p}_i^{ask}(k)\right)$$

Thus, the buyer's value function is a weighted average between the utility of exiting upon a valuation shock, 0, and the net benefit of owning a bond, $V_i^o(k) - \overline{p}_i^{ask}(k)$. The latter is the expected utility as a bond owner minus the cost of becoming an owner in the first place. The relative probabilities of these outcomes determine the relative weights. If the probability of switching and exiting is high, the buyer puts more weight on the value of that outcome. If, instead, the probability of purchasing the bond (i.e. liquidity immediacy) is high, the buyer puts more weight on the net value of owning the bond.

Consider the owner's expected utility:

$$rV_i^o(k) = \delta + k\left(V_i^s - V_i^o(k)\right).$$

From here

$$V_i^o(k) = \frac{\delta + kV_i^s}{r+k}.$$
$$= \frac{r}{r+k}\left(\frac{\delta}{r}\right) + \frac{k}{r+k}\left(V_i^s\right) \tag{B.45}$$

Thus, the owner's expected utility is the weighted average between $\frac{\delta}{r}$ (the present value of the bond coupon flow if one were to hold the bond forever) and V_i^s (the expected utility of a seller). If the probability of getting a valuation shock and, consequently, turning into a seller is high (i.e. k is high), a bond owner puts more weight on what happens to her as a seller, and less on the coupon flow she receives in the meantime.

Finally, consider the seller's expected utility:

$$rV_i^s = \delta - x + \sum_{j\in\{i,N_i\}}\left[\int_{\underline{k}}^{\overline{k}} \lambda_{ij}\hat{\mu}_j^b(k)\left(\hat{p}_{i,j}^{bid}(k) - V_i^s\right)\right]$$
$$= \delta - x + m_i^b\frac{1}{m_i^b}\sum_{j\in\{i,N_i\}}\left[\int_{\underline{k}}^{\overline{k}} \lambda_{ij}\hat{\mu}_j^b(k)\hat{p}_{i,j}^{bid}(k)\right] - V_i^s\sum_{j\in\{i,N_i\}}\left[\int_{\underline{k}}^{\overline{k}} \lambda_{ij}\hat{\mu}_j^b(k)\right]$$
$$= \delta - x + m_i^b\overline{p}_i^{bid} - V_i^s m_i^b$$

From here,

$$
\begin{aligned}
V_i^s &= \frac{\delta - x + m_i^b \overline{p}_i^{bid}}{(r + m_i^b)} \\
&= \frac{r}{(r + m_i^b)} \left(\frac{\delta - x}{r} \right) + \frac{m_i^b}{(r + m_i^b)} \left(\overline{p}_i^{bid} \right). \qquad \text{(B.46)}
\end{aligned}
$$

Thus, the seller's value function is the weighted average between the value of holding the bond forever, $\frac{\delta - x}{r}$, and the expected revenue from selling it, $\overline{p}_i^{bid}$.[40] If the probability of selling, m_i^b, is high, the seller puts more weight on the expected revenue from selling, and less on $\frac{\delta - x}{r}$. The expected utility of the seller, thus, increases with both the price she can sell at, $\overline{p}_i^{bid}$, and the probability of selling it, m_i^b.[41]

Thus, from (B.45), a liquidity investor (i.e. a high k investor) worries more about what happens to her if she is forced to sell later. In particular, from (B.46), she worries about the price at which the dealer buys back the bond from its clients, $\overline{p}_i^{bid}$. Conversely, a buy-and-hold investor cares relatively less about the dealer's bid-price. This is how liquidity investors care more about round trip transaction costs.

To see the benefit of choosing a core dealer specifically, substitute the owner and seller's value functions into the buyer's:

$$
\begin{aligned}
&V_i^b(k) = \\
&= \frac{m_i^s}{r + k + m_i^s} \left(V_i^o(k) - \overline{p}_i^{ask}(k) \right) \\
&= \frac{m_i^s}{r + k + m_i^s} \left[\frac{r}{r + k} \frac{\delta}{r} + \frac{k}{r + k} V_i^s - \overline{p}_i^{ask}(k) \right] \\
&= \frac{m_i^s}{r + k + m_i^s} \left(\frac{r}{r + k} \frac{\delta}{r} + \frac{k}{r + k} \left[\frac{r}{(r + m_i^b)} \frac{\delta - x}{r} + \frac{m_i^b}{(r + m_i^b)} \overline{p}_i^{bid} \right] - \overline{p}_i^{ask}(k) \right) \\
&= \frac{m_i^s}{r + k + m_i^s} \left(\frac{r}{r + k} \frac{\delta}{r} + \frac{k}{(r + k)} \frac{r}{(r + m_i^b)} \frac{\delta - x}{r} - \left[\overline{p}_i^{ask}(k) - \frac{k}{(r + k)} \frac{m_i^b}{(r + m_i^b)} \overline{p}_i^{bid} \right] \right) \\
&= \frac{m_i^s}{r + k + m_i^s} \left(\frac{r}{r + k} \frac{\delta}{r} + \frac{k}{(r + k)} \frac{r}{(r + m_i^b)} \frac{\delta - x}{r} - \phi^{eff}(k) \right),
\end{aligned}
$$

where

$$
\phi_i^{eff} \equiv \overline{p}_i^{ask}(k) - \left[\frac{k}{(r + k)} \frac{m_i^b}{(r + m_i^b)} \right] \left(\overline{p}_i^{bid} \right)
$$

[40]The value of holding the bond forever is simply the present value of the seller's valuation of the bond coupon flow.

[41]Whether V_i^s is increasing in m_i^b, depends on the sign of: $\overline{p}_i^{bid} - \frac{\delta - x}{r}$. If $\overline{p}_i^{bid} - \frac{\delta - x}{r}$ is positive, then V_i^s is increasing in m_i^b also. It must be positive because, intuitively, the seller must be willing to sell only because the expected bid-price is higher than holding the bond forever.

is the effective round trip transaction cost: the expected bid-ask spread scaled by the liquidity immediacy, m_i^b, of the dealer. The effective round trip transaction cost declines (with k) at a faster rate for the clients of a core dealer.[42] Thus, the benefit of choosing a core dealer is that, in relative terms (not necessarily in absolute levels), a core dealer offers a narrower transaction cost. And, as above, prices serve as a sorting device.

Whether the core dealer offers a narrower transaction cost also in absolute levels depends on the liquidity immediacy it offers compared to that of a peripheral dealer. The latter, in turn, depends on λ_{DD} vs λ_D. If $\lambda_{DD} > \lambda_D$, a core dealer offers inferior liquidity immediacy: $m_i^\tau < m_j^\tau$, and, compensating for its inferior liquidity, a core dealer offers a narrower bid-ask spread ($\overline{\phi}_i(k) < \overline{\phi}_j(k)$ for all k). If $\lambda_{DD} = \lambda_D$, core and peripheral dealers offer the same liquidity immediacy, and the point at which a core dealer's bid-ask spread becomes narrower coincides with the endogenous cutoff, k^*. If $\lambda_{DD} < \lambda_D$, a core dealer offers better liquidity immediacy: $m_i^\tau > m_j^\tau$, and a core dealer's bid-ask spread becomes narrower at a point further to right of k^*. In all cases, recall that the core dealer's transaction cost is declining at a faster rate. The intuition is as follows. Given the faster decline of the core dealer's transaction cost, at some k in $[\underline{k}, \overline{k}]$, the two transaction costs cross, and the core dealer's cost becomes lower. The worse the liquidity immediacy of the core dealer, the core dealer's transaction cost has to decline at an even higher rate to compensate for its inferior liquidity immediacy. That is, the benefit of choosing a core dealer has to kick-in sooner (i.e. the point at which they cross shifts to the left). In some cases (as in $\lambda_{DD} > \lambda_D$), it already starts off narrower in absolute levels.

C Proofs

Lemma 2. *For a given cutoff $\{\nu^i(k)\}_i$, the equations characterizing client masses has a unique solution.*

Proof. From buyers' inflow-outflow equation (9),

$$\hat{\mu}_i^b(k) = \frac{\hat{f}(k)\nu_i(k)}{k + \sum\limits_{j\in N} \lambda_{ij}\mu_j^s} \tag{C.47}$$

[42]Similar results hold if we define the effective transaction cost as $\phi_i^{eff} = \left(\frac{k}{(r+k)}\frac{r}{(r+\overline{\mu}_i^b)}\right)\left(\frac{x}{r}\right) + \overline{p}_i^{ask}(k) - \left(\frac{k}{(r+k)}\frac{\overline{\mu}_i^b}{(r+\overline{\mu}_i^b)}\right)\overline{p}_i^{bid}$.

From owners' inflow-outflow equation (A.30) and (C.47),

$$\begin{aligned}\hat{\mu}_i^o(k) &= \frac{\hat{\mu}_i^b(k)\sum\limits_{j\in N}\lambda_{ij}\mu_j^s}{k} \\ &= \frac{\hat{f}(k)\nu_i(k)\sum\limits_{j\in N}\lambda_{ij}\mu_j^s}{k\left(k+\sum\limits_{j\in N}\lambda_{ij}\mu_j^s\right)}\end{aligned} \tag{C.48}$$

Using the market clearing condition (6) and (C.48), the measure of sellers of dealer i, μ_i^s, is determined by:

$$\int_{\underline{k}}^{\bar{k}} \frac{\hat{f}(k)\nu^i(k)\sum\limits_{j\in N}\lambda_{ij}\mu_j^s}{k\left(k+\sum\limits_{j\in N}\lambda_{ij}\mu_j^s\right)}dk + \mu_i^s = s_i \tag{C.49}$$

Using $\sum\limits_{j\in N_i}\mu_j^s = \mu_N^s - \mu_i^s$ and $\sum\limits_{j\in N}\lambda_{ij}\mu_j^s = \lambda_{DD}\mu_N^s - \mu_i^s(\lambda_{DD}-\lambda_D)$, re-express (C.49) as

$$\int_{\underline{k}}^{\bar{k}} \frac{\hat{f}(k)\nu^i(k)\left(\lambda_{DD}\mu_N^s - \mu_i^s(\lambda_{DD}-\lambda_D)\right)}{k\left(k+\lambda_{DD}\mu_N^s - \mu_i^s(\lambda_{DD}-\lambda_D)\right)}dk + \mu_i^s = s_i \tag{C.50}$$

Summing (C.50) across dealers, we get

$$\sum_{i\in N}\left(\int_{\underline{k}}^{\bar{k}} \frac{\hat{f}(k)\nu^i(k)\left(\lambda_{DD}\mu_N^s - (\lambda_{DD}-\lambda_D)\mu_i^s\right)}{k\left(k+\lambda_{DD}\mu_N^s - (\lambda_{DD}-\lambda_D)\mu_i^s\right)}dk\right) + \mu_N^s = S. \tag{C.51}$$

Plugging (C.47) into the interdealer constraint, $\mu_i^s\mu_N^b = \mu_i^b\mu_N^s$, the constraint for each i becomes:

$$\begin{aligned}&\mu_i^s\sum_{i\in N}\left(\int_{\underline{k}}^{\bar{k}} \frac{\hat{f}(k)\nu^i(k)}{k+\lambda_{DD}\mu_N^s-(\lambda_{DD}-\lambda_D)\mu_i^s}dk\right) \\ &= \mu_N^s\int_{\underline{k}}^{\bar{k}} \frac{\hat{f}(k)\nu^i(k)}{k+\lambda_{DD}\mu_N^s-(\lambda_{DD}-\lambda_D)\mu_i^s}dk\end{aligned} \tag{C.52}$$

Thus, $\{\mu_1^s, \mu_2^s, ...\mu_n^s\}$ and μ_N^s is a solution to a system of $n+1$ equations (C.51) and (C.52) for each i. It remains to show that, given a cutoff $\{\nu^i(k)\}_i$, a unique solution exists to (C.51) and (C.52) for each i. $\square$

Lemma 3. *The seller value functions V_i^s and V_j^s for any two dealers i and j in N are given by the solution to* (C.58) *and* (C.59).

Proof. Combining the buyer's and owner's value functions and substituting in prices, we get

$$V_i^{ob}(k) = \frac{\delta + kV_i^s + \sum_{j\in N} u_{ji}\mu_j^s V_j^s}{r + k + \sum_{j\in N} u_{ji}\mu_j^s}.$$

where

$$V_i^{ob}(k) \equiv V_i^o(k) - V_i^b(k)$$

and

$$u_{ji} \equiv \lambda_{ji} z_{ji}$$

Define $\tilde{E}V_i^{ob}(k) \equiv \int_{\underline{k}}^{\overline{k}} \hat{\mu}^b(k) V_i^{ob}(k)$, then

$$\begin{aligned}
&\tilde{E}V_i^{ob}(k) \\
&= \left(\delta + \sum_{j\in N} u_{ji}\mu_j^s V_j^s\right) \int_{\underline{k}}^{\overline{k}} \frac{1}{r + k + \sum_{j\in N} u_{ji}\mu_j^s} \hat{\mu_i}^b(k) dk \\
&+ V_i^s \int_{\underline{k}}^{\overline{k}} \frac{k}{r + k + \sum_{j\in N} u_{ji}\mu_j^s} \hat{\mu_i}^b(k) dk
\end{aligned} \tag{C.53}$$

Defining

$$g_i = \int_{\underline{k}}^{\overline{k}} \frac{1}{r + k + \sum_{j\in N} u_{ji}\mu_j^s} \hat{\mu_i}^b(k) dk \tag{C.54}$$

$$g_i^k = \int_{\underline{k}}^{\overline{k}} \frac{k}{r + k + \sum_{j\in N} u_{ji}\mu_j^s} \hat{\mu_i}^b(k) dk \tag{C.55}$$

Eq.(C.53) becomes

$$\begin{aligned}
\tilde{E}V_i^{ob}(k) &= \left(\delta + \sum_{j\in N} u_{ji}\mu_j^s V_j^s\right) g_i + V_i^s g_i^k \\
&= \left(u\mu_i^s g_i + g_i^k\right) V_i^s + u_I \mu_j^s g_i V_j^s + A_{ij} g_i
\end{aligned}$$

where

$$N_{ij} \equiv N/\{i, j\},$$

$$A_{ij} \equiv \delta + u_I \sum_{j\in N_{ij}} \mu_j^s V_j^s,$$

$$u_I \equiv \lambda_{DD} z_{DD},$$

and

$$u \equiv \lambda_D z_D.$$

Note that $A_{ij} = A_{ji}$.

$$\begin{aligned} u\tilde{E}V_i^{ob}(k) + u_I\tilde{E}V_j^{ob}(k) = &\left[\mu_i^s\left(u^2 g_i + u_I^2 g_j\right) + u g_i^k\right] V_i^s \\ &+ \left[\mu_j^s u u_I g_{ij} + u_I g_j^k\right] V_j^s + A\left(u g_i + u_I g_j\right). \end{aligned} \tag{C.56}$$

where $g_{ij} \equiv g_i + g_j$.

Consider the seller's value function:

$$rV_i^s = \delta - x + \sum_{j\in N}\left(\int_{\underline{k}}^{\bar{k}} u_{ij}\hat{\mu}_j^b(k)\left(V_j^{ob}(k) - V_i^s\right)\right). \tag{C.57}$$

Defining $\mu_{N_{ij}}^b = \sum_{j\in Nij}\mu_j^b$ and $C_{ij} = C_{ji} = \delta - x + u_I\sum_{j\in N_{ij}} EV_j^{ob}(k)$, rewrite (C.57) as

$$\left(r + u\mu_i^b + u_I\mu_j^b + u_I\mu_{N_{ij}}^b\right) V_i^s = u\tilde{E}V_i^{ob}(k) + u_I\tilde{E}V_j^{ob}(k) + C_{ij}.$$

Substituting in (C.56), we get

$$\begin{aligned} &\left(r + u\mu_i^b + u_I\mu_j^b + u_I\mu_{Nij}^b - \mu_i^s\left(u^2 g_i + u_I^2 g_j\right) - u g_i^k\right) V_i^s \\ &= \left(\mu_j^s u u_I\left(gg\right) + u_I g_j^k\right) V_j^s + A_{ij}\left(u g_i + u_I g_j\right) + C_{ij}. \end{aligned} \tag{C.58}$$

The analogous equation characterizing V_j^s is

$$\begin{aligned} &\left(r + u\mu_j^b + u_I\mu_i^b + u_I\mu_{Nij}^b - \mu_j^s\left(u^2 g_j + u_I^2 g_i\right) + u g_j^k\right) V_j^s \\ &= \left(\mu_i^s u u_I\left(gg\right) + u_I g_i^k\right) V_i^s + A_{ij}\left(u g_j + u_I g_i\right) + C_{ij}. \end{aligned} \tag{C.59}$$

Eq.(C.58) and (C.59) characterize V_i^s and V_j^s for any two dealers i and j, irrespective of the aggregate number of dealers. □

Proof of Proposition 1. In a symmetric equilibrium, $\mu_i^s = \frac{\mu_N^s}{n}$ for all i. Using $\sum \nu^i(k) = 1$,(C.51) becomes

$$\int_{\underline{k}}^{\bar{k}} \frac{\hat{f}(k)\left(\lambda_{DD}\mu_N^s - \left(\lambda_{DD} - \lambda_D\right)\frac{\mu_N^s}{n}\right)}{k\left(k + \lambda_{DD}\mu_N^s - \left(\lambda_{DD} - \lambda_D\right)\frac{\mu_N^s}{n}\right)} dk + \mu_N^s - S = 0. \tag{C.60}$$

The left-hand-side is negative at $\mu_N^s = 0$, strictly increasing in μ_N^s, and is equal to ∞ for $\mu_N^s = \infty$. Hence, it has a unique solution. Given the solution for μ_N^s, the other client masses are uniquely determined. Importantly, the solution does not depend on the dealer choice decisions, $\{\nu_i(k)\}_i$. A continuum of dealer choice decisions, as a result, satisfy (C.60) and all of them yield the same solution for client masses.

Consider now the value functions. If $\mu_i^s = \mu_j^s$, the coefficients in front of V_i^s and V_j^s are symmetric across the two equations characterizing V_i^s and V_j^s. As a result, $V_i^s = V_j^s$. Moreover, $V_i^{ob}(k) = V_j^{ob}(k)$. Now consider the buyer's value function:

$$(r+k)\, V_i^b(k) = \sum_{j\in N} u_{ji}\mu_j^s \left(V_i^{ob}(k) - V_j^s\right).$$

Since $V_1^s = V_2^s$, $V_i^{ob}(k) = V_j^{ob}(k)$, and $\mu_i^s = \mu_j^s$, we get that

$$V_i^b(k) = V_j^b(k). \tag{C.61}$$

for all k for any i and j. Thus, the indifference condition is automatically satisfied and does not put an additional constraint on the dealer choice decisions. Moreover, a continuum of dealer choice decisions satisfy (C.61), and, hence, a continuum of symmetric equilibria exist. $\square$

Proof of Lemma 1. Taking the difference between $\hat{V}_2^b(k)$ and $\hat{V}_1^b(k)$ characterized by (C.67), using (C.72), and simplifying, we get

$$\begin{aligned}\hat{V}_2^b(k) - \hat{V}_1^b(k) &= (k-k^*)\,(u_I - u)\,(\mu_2^s - \mu_1^s)\,(\delta - rV_1^s)\,\frac{1}{\hat{p}_2\hat{p}_1 C_0}\cdot \\ &\cdot \left[k\,(k^*+r)\,u_I\mu_N^s + r\,(k^* u_I \mu_N^s + r u_I \mu_N^s + (u_I\mu_1^s + u\mu_2^s)\,(u\mu_1^s + u_I\mu_2^s))\right]\end{aligned}$$

where

$$\begin{aligned}C_0 \equiv{}& k^* u_I \mu_1^s\,(k^* + r + u\mu_1^s) + + k^* u u_I\,(\mu_2^s)^{\,2}, \\ &+ \left((k^*+r)\,(r\,(u_I - u) + k^* u_I) + \left(r\left(u_I^2 - u^2\right) + 2k^* u_I^2\right)\mu_1^s\right)\mu_2^s,\end{aligned}$$

and

$$\hat{p}_i \equiv k + r + u\mu_i^s + u_I\mu_j^s.$$

The entire term multiplying $(k-k^*)\,(u_I-u)\,(\mu_2^s-\mu_1^s)$ is positive. Thus, if $\mu_2^s - \mu_1^s > 0$, $\hat{V}_2^b(k) - \hat{V}_1^b(k) > 0$ for buyers with $k > k^*$; otherwise, $\hat{V}_2^b(k) - \hat{V}_1^b(k) < 0$. That is, it is optimal for buyers with $k > k^*$ to select the dealer with the larger seller client mass, and vice versa for buyers with $k < k^*$. $\square$

Proof of Proposition 2. To simplify notation, we express z_{D} as z and λ_{D}

as λ. We prove existence for the case of two dealers, indexed 1 and 2. In particular, we show that $V_2^b(k^*) - V_1^b(k^*) < 0$ at $k^* = \underline{k}$ and $V_2^b(k^*) - V_1^b(k^*) > 0$ at $k^* = \bar{k}$, which by continuity implies that a cutoff $k^* \in (\underline{k}, \bar{k})$ exists such that $V_2^b(k^*) - V_1^b(k^*) = 0$.

Solving (A.32) for $V_i^o(k)$, we get

$$V_i^o(k) = \frac{\delta + kV_i^s}{k + r}. \tag{C.62}$$

If we set $k^* = \bar{k}$, then $\mu_2^s = 0$ and $\mu_2^b = 0$. Using (C.62) and (A.31), and solving for $V_1^b(k)$ and $V_2^b(k)$, we get:

$$V_1^b(k) = \frac{\mu_1^s u(\delta - rV_1^s)}{(k+r)(k+r+u\mu_1^s)},$$

$$V_2^b(k) = \frac{\mu_1^s u_I(\frac{\delta + kV_2^s}{k+r} - V_1^s)}{(k+r)(k+r+u\mu_1^s)}.$$

Taking the difference $V_2^b(k) - V_1^b(k)$ and multiplying by $\frac{k+r}{\lambda\mu_1^s}$, the sign of $V_2^b(k) - V_1^b(k)$ depends on

$$\begin{aligned} & -\frac{u(\delta - rV_1^s)}{k+r+z\lambda\mu_1^s} + \frac{u_I(\delta - (k+r)V_1^s + kV_2^s)}{k+r+u_I\mu_1^s} && \text{(C.63)} \\ & = -\frac{u(\delta - rV_1^s)}{k+r+z\lambda\mu_1^s} + \frac{u_I(\delta - rV_1^s)}{k+r+u_I\mu_1^s} + \frac{u_I k(V_2^s - V_1^s)}{k+r+u_I\mu_1^s}. && \text{(C.64)} \end{aligned}$$

To determine the sign of (C.63), we first show that $\delta - rV_1^s > 0$ and $\delta - rV_2^s > 0$. Using (A.33), and solving for V_1^s and V_2^s, we get:

$$rV_1^s = \delta - x + x\frac{u\mu_1^b}{k+r+u(\mu_1^b + \mu_1^s)}. \tag{C.65}$$

$$rV_2^s = \delta - x + x\frac{u_I\lambda\mu_1^b(r + u\mu_1^b)}{\left(r + u_I\mu_1^b\right)\left(k+r+u(\mu_1^b + \mu_1^s)\right)}. \tag{C.66}$$

Thus, $rV_1^s = \delta - x(1 - \frac{u\mu_1^b}{k+r+u(\mu_1^b+\mu_1^s)})$, and, hence, $\delta - rV_1^s > 0$. Analogously, $\delta - rV_2^s > 0$.

The term $\frac{u(\delta - rV_1^s)}{k+r+u\mu_1^s}$ is then an increasing function of the term u; thus, $\frac{u_I(\delta - rV_1^s)}{k+r+u_I\mu_1^s} > \frac{u(\delta - rV_1^s)}{k+r+u\mu_1^s}$, and the first two terms (C.63) together are positive. It remains to show that $V_2^s - V_1^s > 0$. The sign of $V_2^s - V_1^s$ depends on the difference of the last terms in (C.65) and (C.66):

$$
\begin{aligned}
&\frac{u_I \mu_1^b (r + u\mu_1^b)}{\left(r + u_I \mu_1^b\right)\left(k + r + u(\mu_1^b + \mu_1^s)\right)} - \frac{u\mu_1^b}{k + r + u(\mu_1^b + \mu_1^s)} \\
&= \frac{\mu_1^b}{k + r + u(\mu_1^b + \mu_1^s)} \frac{u_I - u}{\left(r + u_I \mu_1^b\right)}.
\end{aligned}
$$

Since $u_I - u > 0$, we have $V_2^s - V_1^s > 0$, and consequently $V_2^b(k) - V_1^b(k) > 0$. Thus, as we expand the client base of dealer 1 (hence, shrink the client base of dealer 2) by $k^* \to \bar{k}$, buyers strictly prefer to change their dealer from dealer 1 to dealer 2.

By an analogous argument, if we set $k^* \to \underline{k}$ and expand the client base of dealer 2, while shrinking the client base of dealer 1 to zero, every buyer wants to switch out of dealer 2 and go with dealer 1: $V_2^b(k) - V_1^b(k) < 0$.

Thus, the function $V_2^b(k^*) - V_1^b(k^*)$ is negative at $k^* = \underline{k}$ and positive at $k^* = \bar{k}$. Since it is a continuous function of k^*, there exists k^* such that $V_2^b(k^*) = V_1^b(k^*)$. For any given cutoff, the system of equations has a unique solution. □

Proof of Proposition 3. Suppose $i > j$ (meaning buyers with $k > k^*$ are clients of dealer i). Then, Lemma 4 shows that $\mu_i^s > \mu_j^s$ and $\mu_i^b > \mu_j^b$, Lemma 5 shows that $V_i^s > V_j^s$, and Lemma 6 shows that $\overline{p}_i^{ask}(k) > \overline{p}_j^{ask}(k)$. □

Lemma 4. *If $i > j$, then $\mu_i^s > \mu_j^s$ and $\mu_i^b > \mu_j^b$.*

Proof. The interdealer constraints are

$$\mu_i^s \mu_{N_i}^b = \mu_{N_i}^s \mu_i^b.$$

Substituting in $\mu_{N_i}^b = \mu_N^b - \mu_i^b$ and $\mu_{N_i}^s = \mu_N^s - \mu_i^s$, we get

$$\mu_i^s \left(\mu_N^b - \mu_i^b\right) = \left(\mu_N^s - \mu_i^s\right) \mu_i^b.$$

As a result,

$$\mu_i^b = \mu_i^s \frac{\mu_N^b}{\mu_N^s}.$$

Combining the buyers and owners' value functions, a buyer's value function can be expressed in terms of V^s's:

$$\hat{V}_i^b(k) = \frac{1}{k + r} \frac{u\left(\delta - rV_i^s\right)\mu_i^s + u_I\left(\delta - (k + r)V_j^s + kV_i^s\right)\mu_j^s}{k + r + u\mu_i^s + u_I \mu_j^s}. \tag{C.67}$$

Then, the cutoff k^* where $\hat{V}_2^b(k^*) - \hat{V}_1^b(k^*) = 0$ is given by:

$$\begin{aligned}&\frac{u\left(\delta - rV_2^s\right)\mu_2^s + u_I\left(\delta - (k^*+r)V_1^s + k^*V_2^s\right)\mu_1^s}{k^* + r + u\mu_2^s + u_I\mu_1^s} \\ &- \frac{u\left(\delta - rV_1^s\right)\mu_1^s + u_I\left(\delta - (k^*+r)V_2^s + k^*V_1^s\right)\mu_2^s}{k^* + r + u\mu_1^s + u_I\mu_2^s} = 0.\end{aligned} \tag{C.68}$$

Solving for V_1^s and V_2^s from the linear system given by (C.58)-(C.59), substituting the solution into (C.68), and simplifying, (C.68) becomes

$$\begin{aligned}&(k^*+r)\left(1 + u_I\left(g_1^k + g_2^k\right)\right)\left(\mu_1^s - \mu_2^s\right) \\ &+ \Big[(k^*)^2 u_I\left(\mu_1^s + \mu_2^s\right) + r\left(r + u\mu_1^s + u_I\mu_2^s\right)\left(-u\mu_2^s + u_I\mu_1^s\right) \\ &+ 2k^*u_I\mu_1^s\left(r + u\mu_1^s\right) + k^*\left(r\left(u_I - u\right) + 2u_I^2\mu_1^s\right)\mu_2^s\Big] g_1 \\ &- \Big[(k^*)^2 u_I\left(\mu_1^s + \mu_2^s\right) + r\left(r + u\mu_2^s + u_I\mu_1^s\right)\left(-u\mu_1^s + u_I\mu_2^s\right) \\ &+ k^*\left(r\left(u_I - u\right)\mu_1^s + 2u_I\left(r + u_I\mu_1^s\right)\mu_2^s + 2uu_I\left(\mu_2^s\right)^2\right)\Big] g_2 \\ &= 0.\end{aligned} \tag{C.69}$$

To simplify this further, consider μ_i^b versus g_i^k.

$$\begin{aligned}\frac{1}{r}\int_{\underline{k}}^{k^*} \hat{\mu}_1^b\,dk &= \frac{1}{r}\int_{\underline{k}}^{k^*} \frac{k + r + u\mu_1^s + u_I\mu_2^s}{k + r + u\mu_1^s + u_I\mu_2^s}\hat{\mu}_1^b dk \\ &= \frac{1}{r}\int_{\underline{k}}^{k^*}\left(\frac{k}{k + r + u\mu_1^s + u_I\mu_2^s} + \frac{r + u\mu_1^s + u_I\mu_2^s}{k + r + u\mu_1^s + u_I\mu_2^s}\right)\hat{\mu}_1^b dk \\ &= g_1^k + \left(r + u\mu_1^s + u_I\mu_2^s\right) g_1.\end{aligned}$$

Thus,

$$g_i^k = \frac{1}{r}\mu_i^b - \left(r + u\mu_i^s + u_I\mu_j^s\right) g_i. \tag{C.70}$$

Substituting (C.70) back into (C.69), grouping the terms multiplying g_1 and g_2, and simplifying, we get

$$\begin{aligned}&\left(\mu_2^s - \mu_1^s\right)\left(1 + \frac{1}{r}u_I\mu_N^b\right) + \left[r\left(u_I - u\right)\mu_1^s + k^*u_I\mu_N^s\right]\frac{1}{r + k^*}p_2 g_2 \\ &- \left[r\left(u_I - u\right)\mu_2^s + k^*u_I\mu_N^s\right]\frac{1}{r + k^*}p_1 g_1 = 0.\end{aligned} \tag{C.71}$$

where

$$p_i \equiv k^* + r + u\mu_i^s + u_I\mu_j^s.$$

Now consider $p_i g_i$ versus μ_i^b:

$$\begin{aligned}
&\frac{1}{r}\int_{\underline{k}}^{k^*}\frac{k^*+r+u\mu_1^s+u_I\mu_2^s}{k+r+u\mu_1^s+u_I\mu_2^s}\hat{\mu}_1^b dk-\frac{1}{r}\int_{\underline{k}}^{k^*}\hat{\mu}_1^b dk\\
&=\frac{1}{r}\int_{\underline{k}}^{k^*}\frac{k^*+r+u\mu_1^s+u_I\mu_2^s}{k+r+u\mu_1^s+u_I\mu_2^s}\hat{\mu}_1^b dk-\frac{1}{r}\int_{\underline{k}}^{k^*}\frac{k+r+u\mu_1^s+u_I\mu_2^s}{k+r+u\mu_1^s+u_I\mu_2^s}\hat{\mu}_1^b dk\\
&=\frac{1}{r}\int_{\underline{k}}^{k^*}\frac{k^*-k}{k+r+u\mu_1^s+u_I\mu_2^s}\hat{\mu}_1^b dk>0.
\end{aligned}$$

Thus, $p_1 g_1 > \frac{1}{r}\mu_1^b$ or $-p_1 g_1 < -\frac{1}{r}\mu_1^b$. Analogously, $p_2 g_2 < \frac{1}{r}\mu_2^b$.

The inequalities, $-p_1 g_1 < -\frac{1}{r}\mu_1^b$ and $p_2 g_2 < \frac{1}{r}\mu_2^b$, and (C.71) together imply

$$\begin{aligned}
&(\mu_2^s-\mu_1^s)\left(1+\frac{1}{r}u_I\mu_N^b\right)+\left(r\left(u_I-u\right)\mu_1^s+k^*u_I\mu_N^s\right)\frac{1}{r+k^*}\frac{1}{r}\mu_2^b\\
&-\left(r\left(u_I-u\right)\mu_2^s+k^*u_I\mu_N^s\right)\frac{1}{r+k^*}\frac{1}{r}\mu_1^b>0.
\end{aligned}$$

Using the fact that $\mu_1^b\mu_2^s=\mu_1^s\mu_2^b$ from the interdealer constraint, this becomes

$$(\mu_2^s-\mu_1^s)\left(1+\frac{1}{r}u_I\mu_N^b\right)+\left(k^*u_I\mu_N^s\right)\frac{1}{r+k^*}\frac{1}{r}\mu_2^b-\left(k^*u_I\mu_N^s\right)\frac{1}{r+k^*}\frac{1}{r}\mu_1^b>0,$$

or

$$(\mu_2^s-\mu_1^s)\left(1+\frac{1}{r}u_I\mu_N^b\right)+\frac{1}{r+k^*}\frac{1}{r}\left(k^*u_I\mu_N^s\right)\left(\mu_2^b-\mu_1^b\right)>0.$$

Using the interdealer constraint, $\mu_i^b=\mu_i^s\frac{\mu_N^b}{\mu_N^s}$, one more time, we get

$$(\mu_2^s-\mu_1^s)\left(1+\frac{1}{r}u_I\mu_N^b\right)+\frac{1}{r+k^*}\frac{1}{r}k^*u_I\left(\mu_2^s-\mu_1^s\right)\mu_N^b>0.$$

The terms multiplying $\mu_2^s-\mu_1^s$ are positive. Thus, it has to be that

$$\mu_2^s-\mu_1^s>0.$$

In turn, since from the interdealer constraint $\mu_i^b=\mu_i^s\frac{\mu_N^b}{\mu_N^s}$, we have:$\mu_2^b-\mu_1^b>0$. □

Lemma 5. *If $i>j$, then $V_i^s>V_j^s$.*

Proof. Solving for V_2^s from (C.68), we get

$$V_2^s = \frac{1}{c}(k^* + r)(u_I - u)\,\delta\,(\mu_2^s - \mu_1^s) \tag{C.72}$$
$$+ \frac{1}{c}V_1^s\Big[(k^*)^2 u_I(\mu_1^s + \mu_2^s) + r(u_I - u)\mu_1^s(r + (u + u_I)\mu_2^s)$$
$$+ k^*\left(r\left((u_I - u)\mu_1^s + u_I(\mu_1^s + \mu_2^s)\right) + u_I\left(2u_I\mu_1^s\mu_2^s + u\left((\mu_1^s)^2 + (\mu_2^s)^2\right)\right)\right)\Big].$$

where

$$c \equiv k^* u_I \mu_1^s (k^* + r + u\mu_1^s) + k^* u u_I (\mu_2^s)^2$$
$$+ \left[(k^* + r)(r(u_I - u) + k^* u_I) + \left(r(u_I^2 - u^2) + 2k^* u_I^2\right)\mu_1^s\right]\mu_2^s.$$

Note that $c > 0$. Subtracting V_1^s from (C.72) and simplying, we get:

$$V_2^s - V_1^s = \frac{(k^* + r)(u_I - u)(\delta - rV_1^s)(\mu_2^s - \mu_1^s)}{c}.$$

By assumption, $u_I - u > 0$, and, as the previous lemma showed, $\mu_2^s - \mu_1^s > 0$. Thus, $V_2^s - V_1^s > 0$. □

Lemma 6. *If $i > j$, then $\overline{p}_i^{ask}(k) > \overline{p}_j^{ask}(k)$.*

Proof. To be completed. □

Proof of Proposition 4. Given that dealer i has a larger buyer and seller client mass, it is straightforward to see that $M_i^D > M_j^D$.

Now consider the interdealer volume, M_i^{DD}. Due to the interdealer constraints, the first and the second terms in (5) are equal. Hence, M_i^{DD} is the twice the first term:

$$M_i^{DD} = 2\lambda_{\text{DD}}\mu_i^s\left(\sum_{j\in N_i}\mu_j^b\right).$$

Using the fact that $\mu_i^b = \mu_i^s\frac{\mu_N^b}{\mu_N^s}$,

$$M_i^{DD} = 2\lambda_{\text{DD}}\mu_i^s\left(\sum_{j\in N_i}\mu_j^s\right)\frac{\mu_N^b}{\mu_N^s}. \tag{C.73}$$

Thus, the sign of $M_i^{DD} - M_j^{DD}$ depends on the sign of:

$$\begin{aligned}
&\mu_i^s \mu_{N_i}^s - \mu_j^s \mu_{N_j}^s \\
&= \mu_i^s \left(\mu_j^s + \mu_{N/\{i,j\}}^s\right) - \mu_j^s \left(\mu_i^s + \mu_{N/\{i,j\}}^s\right) \\
&= \left(\mu_i^s - \mu_j^s\right) \mu_{N/\{i,j\}}^s.
\end{aligned}$$

As a result, $M_i^{DD} - M_j^{DD} > 0$ because $\mu_i^s - \mu_j^s > 0$. □

Proof of Proposition 5. Using (A.35)

$$\begin{aligned}
P_{i,d}^{ask} \equiv E_d^b \left[\hat{P}_{i,d}(k)\right] &= \frac{1}{2} V_i^s \int_{\underline{k}}^{\overline{k}} \frac{\hat{\mu}_j^b(k)}{\mu_j^b} + \frac{1}{2} \int_{\underline{k}}^{\overline{k}} \frac{\hat{\mu}_d^b(k)}{\mu_d^b} (V_d^o(k) - V_d^b(k)). \\
&= \frac{1}{2} V_i^s + \frac{1}{2} E_d^b \left[V_d^{ob}(k)\right]. \qquad \text{(C.74)}
\end{aligned}$$

Since $V_c^s > V_p^s$,

$$P_{c,d}^{ask} > P_{p,d}^{ask}.$$

Thus, a dealer buys an asset at a higher ask-price from a core dealer than from a peripheral dealer.

Now consider the price an abritrary dealer d sells back to dealer i

$$\begin{aligned}
P_{d,i}^{bid} &= \frac{1}{2} V_d^s + \frac{1}{2} \int_{\underline{k}}^{\overline{k}} \frac{\hat{\mu}_i^b(k)}{\mu_b^b} V_i^{ob}(k). \qquad \text{(C.75)} \\
&= \frac{1}{2} V_d^s + \frac{1}{2} E_i^b V_i^{ob}. \qquad \text{(C.76)}
\end{aligned}$$

Since core dealer's clients are high k-buyers, and high k-buyers have low reservation values, $E_c^b \left[V_c^o(k) - V_c^b(k)\right] < E_p^b \left[V_p^o(k) - V_p^b(k)\right]$. Thus,

$$P_{d,c}^{bid} < P_{d,p}^{bid}$$

Combining the two, the core dealer charges a wider bid-ask spread:

$$\frac{P_c^{ask} - P_c^{bid}}{0.5 P_c^{ask} + 0.5 P_c^{bid}} > \frac{P_p^{ask} - P_p^{bid}}{0.5 P_p^{ask} + 0.5 P_p^{bid}}.$$

The results on execution speed and volume are straightforward implications from the difference in client sizes across dealers. □

Proof of Proposition 6. Integrating the value functions over the respective client masses yields:

$$r\int_{\underline{k}}^{\overline{k}} V_i^o(k)\hat{\mu}_i^o(k)dk = \delta\int_{\underline{k}}^{\overline{k}} \hat{\mu}_i^o(k)dk + k\int_{\underline{k}}^{\overline{k}} \left(V_i^s - V_i^o(k)\right)\hat{\mu}_i^o(k)dk.$$

$$\begin{aligned} r\int_{\underline{k}}^{\overline{k}} V_i^b(k)\hat{\mu}_i^b(k) = &\int_{\underline{k}}^{\overline{k}} k\left(0 - V_i^b(k)\right)\hat{\mu}_i^b(k)dk \\ &+ \int_{\underline{k}}^{\overline{k}} \sum_{j\in N}\lambda_{ij}\mu_j^s z_{ij}\left(V_i^o(k) - V_i^b(k) - V_j^s\right)\hat{\mu}_i^b(k)dk. \end{aligned}$$

$$rV_i^s\mu_i^s = (\delta - x)\,\mu_i^s + \sum_{j\in N}\left(\int_{\underline{k}}^{\overline{k}} \lambda_{ij}\mu_i^s\hat{\mu}_j^b(k)z_{ij}\left(V_j^o(k) - V_j^b(k) - V_i^s\right)\right).$$

Adding these up, plus the new entrants expected utility $\int_{\underline{k}}^{\overline{k}} V_i^b(k)\hat{f}(k)\nu_i(k)dk$ and dealer profits rW_i^D, we get

$$\begin{aligned} r(W_i^C + W_i^D) =&\delta\int_{\underline{k}}^{\overline{k}} \hat{\mu}_i^o(k)dk + \int_{\underline{k}}^{\overline{k}} k\left(V_i^s - V_i^o(k)\right)\hat{\mu}_i^o(k)dk \\ &+ \int_{\underline{k}}^{\overline{k}} k\left(0 - V_i^b(k)\right)\hat{\mu}_i^b(k)dk \\ &+ \int_{\underline{k}}^{\overline{k}} \sum_{j\in N}\lambda_{ij}\mu_j^s z_{ij}\left(V_i^o(k) - V_i^b(k) - V_j^s\right)\hat{\mu}_i^b(k)dk \\ &+ (\delta - x)\,\mu_i^s + \sum_{j\in N}\left(\int_{\underline{k}}^{\overline{k}} \lambda_{ij}\mu_i^s\hat{\mu}_j^b(k)z_{ij}\left(V_j^o(k) - V_j^b(k) - V_i^s\right)\right) \\ &+ \int_{\underline{k}}^{\overline{k}} V_i^b(k)\hat{f}(k)\nu_i(k)dk \\ &+ \int_{\underline{k}}^{\overline{k}} \lambda_{ij}\hat{\mu}_i^b(k)\mu_i^s\,(1 - 2z)\left(V_i^o(k) - V_i^b(k) - V_i^s\right)dk \\ &+ \sum_{j\in N_i}\left(\int_{\underline{k}}^{\overline{k}} \lambda_{ij}\hat{\mu}_i^b(k)\mu_j^s\left(\frac{1 - 2z_{\mathrm{DD}}}{2}\right)\left(V_i^o(k) - V_i^b(k) - V_j^s\right)dk\right) \\ &+ \sum_{j\in N_i}\left(\int_{\underline{k}}^{\overline{k}} \lambda_{ij}\hat{\mu}_j^b(k)\mu_i^s\left(\frac{1 - 2z_{\mathrm{DD}}}{2}\right)\left(V_j^o(k) - V_j^b(k) - V_i^s\right)dk\right). \end{aligned}$$

Simplifying it and replacing $\hat{\mu}_i^b(k)$ and $\hat{\mu}_i^o(k)$ with $\hat{\mu}_i^b(k) = \frac{\hat{f}(k)\nu_i(k)}{k+\lambda\mu_{iN}^s}$ and

$\hat{\mu}_i^o(k) = \frac{\hat{f}(k)\nu_i(k)\lambda\mu_{iN}^s}{k\left(k+\lambda\mu_{iN}^s\right)}$, we get

$$
\begin{aligned}
r(W_i^C + W_i^D) =& \delta \int_{\underline{k}}^{\overline{k}} \frac{\hat{f}(k)\nu_i(k)\lambda\mu_{iN}^s}{k\left(k+\lambda\mu_{iN}^s\right)} dk + \int_{\underline{k}}^{\overline{k}} \left(V_i^s - V_i^o(k)\right) \frac{\hat{f}(k)\nu_i(k)\lambda\mu_{iN}^s}{\left(k+\lambda\mu_{iN}^s\right)} dk. \\
&+ \int_{\underline{k}}^{\overline{k}} k\left(0 - V_i^b(k)\right) \frac{\hat{f}(k)\nu_i(k)}{k+\lambda\mu_{iN}^s} dk \\
&+ \int_{\underline{k}}^{\overline{k}} \lambda\mu_i^s \left(V_i^o(k) - V_i^b(k) - V_i^s\right) \hat{\mu}_i^b(k) dk \\
&+ \int_{\underline{k}}^{\overline{k}} \sum_{j\in N_i} \lambda_{ij}\mu_j^s(\frac{1}{2}) \left(V_i^o(k) - V_i^b(k) - V_j^s\right) \hat{\mu}_i^b(k) dk \\
&+ (\delta - x)\mu_i^s + \sum_{j\in N_i} \left(\int_{\underline{k}}^{\overline{k}} \lambda_{ij}\mu_i^s \hat{\mu}_j^b(k)(\frac{1}{2}) \left(V_j^o(k) - V_j^b(k) - V_i^s\right) \right) \\
&+ \int_{\underline{k}}^{\overline{k}} V_i^b(k)\hat{f}(k)\nu_i(k) dk.
\end{aligned}
$$

Adding the second term in the first row, the first term in the second row and the very last term, we get

$$
\begin{aligned}
r(W_i^C + W_i^D) =& \delta \int_{\underline{k}}^{\overline{k}} \frac{\hat{f}(k)\nu_i(k)\lambda\mu_{iN}^s}{k\left(k+\lambda\mu_{iN}^s\right)} dk - \lambda\mu_{iN}^s \int_{\underline{k}}^{\overline{k}} \left(V_i^o(k) - V_i^b(k) - V_i^s\right) \hat{\mu}_i^b(k) dk. \\
&+ \lambda\mu_i^s \int_{\underline{k}}^{\overline{k}} \left(V_i^o(k) - V_i^b(k) - V_i^s\right) \frac{\hat{f}(k)\nu_i(k)}{k+\lambda\mu_{iN}^s} dk \\
&+ \int_{\underline{k}}^{\overline{k}} \sum_{j\in N_i} \lambda_{ij}\mu_j^s(\frac{1}{2}) \left(V_i^o(k) - V_i^b(k) - V_j^s\right) \hat{\mu}_i^b(k) dk \\
&+ (\delta - x)\mu_i^s + \sum_{j\in N_i} \left(\int_{\underline{k}}^{\overline{k}} \lambda_{ij}\mu_i^s \hat{\mu}_j^b(k)(\frac{1}{2}) \left(V_j^o(k) - V_j^b(k) - V_i^s\right) \right).
\end{aligned}
$$

Summing across all dealers $i \in N$ and using the fact $\mu_i^b = \mu_i^s \frac{\mu_N^b}{\mu_N^s}$, all the expressions involving V's cancel. We are left with:

$$
\begin{aligned}
&\sum_{i\in N} \left(\delta \int_{\underline{k}}^{\overline{k}} \frac{\hat{f}(k)\nu_i(k)\lambda\mu_{i,N_i}^s}{k\left(k+\lambda\mu_{i,N_i}^s\right)} dk + (\delta - x)\mu_i^s \right) \\
&= \sum_{i\in N} \left(\delta(s_i - \mu_i^s) + (\delta - x)\mu_i^s\right) \\
&= \delta S - x\mu_N^s,
\end{aligned}
$$

where the second equality comes from the market clearing condition. $\square$

Proof of Proposition 7. To be completed. $\square$

Proof of Proposition 8. The result is shown in Lemma 4. $\square$

Proof of Proposition 9. Consider an environment with two dealers i and j. Writing $\mu_j^s = \mu_N^s - \mu_i^s$, (C.51) and (C.52) boil down to two equations and two unknowns, μ_i^s and μ_N^s. Using the Implicit Function Theorem, $\frac{\partial \mu_N^s}{\partial k^*}$ evaluated at $k^* = \underline{k}$ (that is, $\mu_i^s = 0$) is

$$\frac{\partial \mu_N^s(k^*)}{\partial k^*} = \frac{f\lambda\mu_N^s(2(\bar{k} - \underline{k})\lambda\mu_N^s + \underline{k}(S - \mu_N^s)(\bar{k} + \lambda\mu_N^s))}{\underline{k}(\underline{k} + \lambda_{\text{DD}}\mu_N^s)\left[\lambda + (\underline{k} + \lambda\mu_N^s)(\bar{k} + \lambda\mu_N^s)\right]\left[-(S - \mu_N^s)\right]}.$$

The numerator is positive, while the denominator is negative; hence, $\frac{\partial \mu_N^s(k^*)}{\partial k^*} < 0$. This implies that as we go from an environment with just one dealer ($k^* = \underline{k}$) to an environment with two dealers ($\underline{k} < k^* < \bar{k}$) (that is, as k^* increases), the misallocation—captured by μ_N^s —decreases. The social welfare, as a result, increases. Thus, increasing the aggregate number of dealers increases the social welfare. $\square$

Proofs of Propositions 10-13. To be completed.

□

D Tables

Table 1: Parameter Values

This table gives the parameter values chosen for the numerical analysis. We assume a uniform distribution for $f(k)$.

Variable	Notation	Value
Bond coupon blow	δ	1
Disutility of holding the bond	x	0.5
Support of customer distribution	$[\underline{k}, \bar{k}]$	[1,5]
Dealers' matching efficiency, CDC	λ_{D}	100
Dealers' matching efficiency, CDDC	λ_{DD}	200
Supply of bonds	S	0.3
Risk-free rate	r	0.04
Customer bargaining power, n=1	z_{D}	$\frac{1}{4}$
Customer bargaining power, n=2	z_{DD}	$\frac{1}{4}$

E Model Figures

Figure 7: The Average Ask-Price vs. the Sellers' Expected Utility

The figures illustrate the tradeoff that buyers face when choosing dealers, for exposition, in a two-dealer environment. They plot the the average ask-price and the expected utility of a seller-client that dealers offer to clients as functions of clients' liquidity type k (in x-axis). The cutoff k^* is the equilibrium cutoff: Buyers with $k < k^*$ choose the "peripheral" dealer, while buyers with $k \geq k^*$ choose the "core" dealer. See Section 2 for more detail.

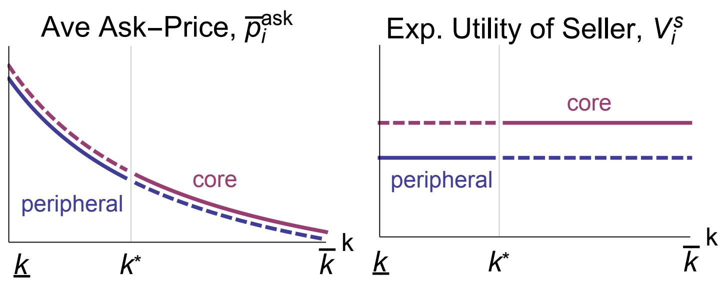


Figure 8: The Quoted Average Bid-Ask Spread

The figures illustrate the expected bid-ask spread, $\overline{\phi}_i(k)$, as functions of buyers' liquidity type k (in x-axis) for three different parameter regions: $\lambda_{DD} < \lambda_D$, $\lambda_{DD} = \lambda_D$, $\lambda_{DD} > \lambda_D$. The cutoff k^* is the equilibrium cutoff: Buyers with $k < k^*$ choose the "peripheral" dealer, while buyers with $k \geq k^*$ choose the "core" dealer. See Section 3 for more detail.

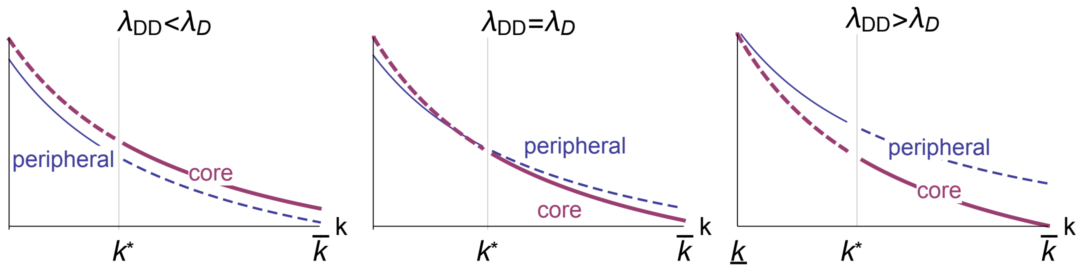
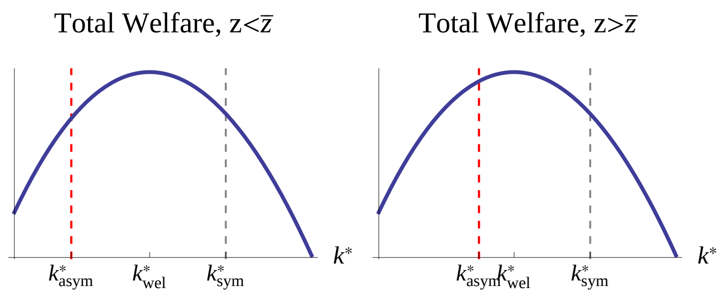


Figure 9: Welfare Analysis

The figures plot, for a two-dealer environment, the total welfare as a function of the cutoff k^*. The cutoff k^*_{asym} is the actual (asymmetric) equilibrium cutoff, k^*_{sym} is a hypothetical cutoff such that $\mu^s_1 = \mu^s_2$, and k^*_{wel} is a cutoff that maximizes the total welfare. we assume that condition 11 is satisfied via $\lambda_{\mathrm{DD}} > \lambda_{\mathrm{D}}$. See Section 4 for more detail.



References

Afonso, Gara, A Kovner, and A Schoar, 2013, Trading partners in the interbank lending market, *FRB of New York Staff Report.*

Atkeson, Andrew G, Andrea L Eisfeldt, and Pierre-Olivier B Weill, 2014, Entry and exit in OTC derivatives markets, Working paper.

Bech, M L, and E Atalay, 2010, The topology of the federal funds market, *Physica A: Statistical Mechanics and its Applications* 389, 5223–5246.

Chang, Briana, and Shengxing Zhang, 2015, Endogenous market making and network formation, Working paper.

Colliard, Jean-Edouard, and Gabrielle Demange, 2014, Cash providers: asset dissemination over intermediation chains, Working paper.

Duffie, Darrell, N Garleanu, and Lasse Heje Pedersen, 2005, Over-the-counter markets, *Econometrica* 73, 1815–1847.

Duffie, Darrell, Semyon Malamud, and Gustavo Manso, 2009, Information percolation with equilibrium search dynamics, *Econometrica.*

Farboodi, Maryam, 2014, Intermediation and voluntary exposure to counterparty risk, Working paper.

Glode, Vincent, and Christian C Opp, 2014, Adverse selection and intermediation chains, Working paper.

Gofman, Michael, 2011, A network-based analysis of over-the-counter markets, Working paper.

Hugonnier, J, B Lester, and Pierre-Olivier B Weill, 2014, Heterogeneity in decentralized asset markets, Working paper.

Kondor, P, and A Babus, 2013, Trading and information diffusion in over-the-counter markets, Working paper.

Lagos, Ricardo, and Guillaume Rocheteau, 2009, Liquidity in asset markets with search frictions, *Econometrica* 77, 403–426.

Li, Dan, and Norman Schürhoff, 2014, Dealer networks: market quality in over-the-counter markets, Working paper.

Malamud, Semyon, and Marzena J Rostek, 2014, Decentralized exchange, Working paper.

Neklyudov, A, 2012, Bid-ask spreads and the over-the-counter interdealer markets: Core and peripheral dealers, Working paper.

Neklyudov, Artem, Burton Hollifield, and Chester Spatt, 2014, Bid-ask spreads, trading networks and the pricing of securitizations: 144a vs. registered securitizations, Working paper.

Sambalaibat, Batchimeg, 2014, A Theory of Liquidity Spillover Between Bond and CDS Markets, Working paper.

Shen, Ji, Bin Wei, and Hongjun Yan, 2015, Financial intermediation chains in a search market, Working paper.

Vayanos, Dimitri, and Tan Wang, 2007, Search and endogenous concentration of liquidity in asset markets, *Journal of Economic Theory* 136, 66–104.

Vayanos, Dimitri, and Pierre-Olivier B Weill, 2008, A search-based theory of the on-the-run phenomenon, *The Journal of Finance* 63, 1361–1398.

Viswanathan, S, and James J D Wang, 2004, Inter-dealer trading in financial markets, *The Journal of Business* 77, 987–1040.

Wang, Jessie Jiaxu, 2014, Distress dispersion and systemic risk in networks, Working paper.

Weill, Pierre-Olivier B, 2008, Liquidity premia in dynamic bargaining markets, *Journal of Economic Theory* 140, 66–96.

Zhong, Zhuo, 2014, The risk sharing benefit versus the collateral cost: The formation of the inter-dealer network in over-the-counter trading, Working paper.