FTG Working Paper Series

Endogenous Specialization and Dealer Networks

by

Batchimeg Sambalaibat

Working Paper No. 00014-02

Finance Theory Group

www.financetheory.com

Endogenous Specialization and Dealer Networks*

Batchimeg Sambalaibat†

July 30, 2020

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

OTC markets exhibit a core-periphery interdealer 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 directed search 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, directed search, clientele effect, specialization, intermediation chains, over-the-counter (OTC) markets, search frictions, interdealer markets.

*I thank Daniel Andrei, Jennie Bai (discussant), Jiakai Chen (discussant), Briana Chang (discussant), Maryam Farboodi (discussant), Song Han (discussant), Barney Hartman-Glaser, Bernard Herskovic (discussant), Burton Hollifield, Ben Lester (discussant), Ioanid Rosu (discussant), Norman Schuerhoff, Chris Telmer, Semih Uslu (discussant), Colin Ward, Pierre-Olivier Weill, Randy Wright, Xiaoyan Zhang (discussant), Ariel Zetlin-Jones, seminar participants at UCLA Anderson, University of Michigan, Vanderbilt University, Federal Reserve Bank of New York, University of Wisconsin Madison, Indiana University, Stockholm School of Economics, Copenhagen Business School, IE Business School, the SEC, conference participants at the AFA, Utah Winter Finance, Stanford Institute for Theoretical Economics, ASU Sonoran Winter Finance Conference, Society of Economic Dynamics, ESSFM Gerzensee, LAEF OTC Markets Workshop, 4th LSE Economic Networks & Finance Conference, EFA, FIRS, MFA, NFA, FTG Summer (Madrid), Early Career Women in Finance, 8th Erasmus Liquidity Conference, Wabash River Conference, Summer Workshop on Money, Banking, Payments & Finance (St. Louis Fed), Southwest Search and Matching (USC), and 11th Annual Central Bank Workshop on the Microstructure of Financial Markets. I especially thank Artem Neklyudov for many helpful conversations.

†Bendheim Center for Finance, Princeton University and Kelley School of Business, Indiana University (sbchimeg@gmail.com)

1 Introduction

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 clients form around dealers determines the shape of the

[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)).

interdealer network. This insight is the main contribution of the paper.

We formalize this insight with a directed search model that builds on Duffie, Gârleanu, 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 liquidity service. Some dealers charge a high ask-price but, in return, offer better liquidity 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 worse liquidity. 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 liquidity the dealer offers 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,

[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 an intermediate asset valuation resemble core dealers, while agents

the 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 2 presents the model. In Section 3, we derive the asymmetric specialization equilibrium and show that the dealer network has a core-periphery structure. Section 4 compares liquidity and prices that core and peripheral dealers provide to customers and,

---

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.

on the interdealer market, to other dealers. Section 5 derives additional results on dealer interconnectedness, market fragmentation, and welfare. In Section 6, we discuss our assumptions. Section 7 concludes.

1.1 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, Gârleanu, 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), 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

[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).

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)).

2 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 n ex-ante identical dealers. Dealers are indexed by $i \in N$, where $N = \{1, 2, 3, \ldots n\}$ is the set of dealers.[8] Everyone is risk neutral, infinitely lived, and discounts the future at a constant rate $r > 0$.

2.1 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 state is an absorbing 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

[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.

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.

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

[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.

and owners with switching rates k in $[k, k+dk]$.

2.2 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 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.

We measure a dealer's network centrality by its volume of interdealer trades, M_i^{DD}, given

[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.

[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.

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

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 even though 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.

2.3 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:

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

2.4 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.

2.5 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 $\lambda_{ij} = \lambda_{\text{DD}}$ if $i \neq j$; otherwise, $\lambda_{ij} = \lambda_{\text{D}}$. 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 C analogously characterizes the owner and seller masses.

2.6 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 $z_{ij} = z_{\text{DD}}$ if $i \neq j$ (i.e. 2-dealer chain); otherwise, $z_{ij} = z_{\text{D}}$. 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 (C.7), who turns around and sells to dealer j at the interdealer price $\hat{P}_{i,j}(k)$ in (C.8). Dealer j, in turn, sells to its buyer-client at the ask price $\hat{p}_{i,j}^{ask}(k)$ in (C.9). 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 C characterizes these expected prices and the expected bid-ask spread $\overline{\phi}_i(k)$ customers face from their dealers. Equations (17)-(18) characterize the probability of getting matched with a buyer, m_i^b, and a seller, m_i^s, respectively.

We make the following assumption on the bargaining powers and matching efficiencies. It is a necessary assumption for dealer heterogeneity to arise.

Assumption 1.

$$\lambda_{\text{DD}} z_{\text{DD}} > \lambda_{\text{D}} z_{\text{D}}. \tag{10}$$

2.7 Continuation Values

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:

$$\begin{aligned} V_i^b(k) = (1 - rdt)\Bigg(& kdt0 + \sum_{j \in N} \lambda_{ij} \mu_j^s dt (V_i^o(k) - \hat{p}_{j,i}^{ask}(k)) + \\ & + [1 - kdt - \sum_{j \in N} \lambda_{ij} \mu_j^s dt] V_i^b(k) \Bigg). \end{aligned} \tag{11}$$

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

The continuation value of a seller client, V_i^s, summarizes the quality of a dealer's liquidity service and, as a result, plays an important role later. The Appendix shows that we can express V_i^s as

$$V_i^s = \frac{r}{(r + m_i^b)} \left(\frac{\delta - x}{r} \right) + \frac{m_i^b}{(r + m_i^b)} \left(p_i^{bid} \right), \tag{12}$$

where m_i^b is the rate at which dealer i executes sell orders and $\overline{p}_i^{bid}$ is the average bid-price of dealer i. From (12), V_i^s is a weighted average between the value of holding the bond forever, $\frac{\delta - x}{r}$, and the expected revenue from selling it, $\overline{p}_i^{bid}$. If the probability of selling (m_i^b) is high, the seller puts more weight on the latter than on $\frac{\delta - x}{r}$. Since $\overline{p}_i^{bid} > \frac{\delta - x}{r}$ in equilibrium, the seller's continuation value increases in both the price she expects to sell back to the dealer ($\overline{p}_i^{bid}$) and the dealer's execution speed (m_i^b). Thus, a dealer offers better liquidity than another dealer if she either buys back at a higher bid-price, executes orders more quickly, or both.

2.8 Equilibrium Characterization

Our analysis focuses on the steady state equilibrium. It is expected utilities $\left\{V_i^o(k), V_i^b(k), V_i^s\right\}_{i\in N}$, population measures $\left\{\hat{\mu}_i^o(k), \hat{\mu}_i^b(k), \mu_i^s\right\}_{i\in N}$, 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 dealer choices $\{\nu_i(k)\}_{i\in N}$ such that (i) value functions solve investors' optimization problems (C.4)–(C.6); (ii) population measures and the distribution of bonds across dealers solve inflow-outflow equations (9), (C.1), market clearing conditions (6)–(7), and inter-dealer transactions equations (8), (iii) prices arise from bargaining (C.7)–(C.9), and (iv) entry decisions solve (1) and $\sum_{i\in N}\nu_i(k) = 1$.

3 Main Results

To characterize the types of equilibria that arise in the model and the client structure in each equilibrium, Lemma 1 establishes a sorting mechanism of clients into dealers. According to Lemma 1, the quality of a dealer's liquidity service—captured by the value function of a seller client (V_i^s)—arises as a sorting device. If dealer i provides better liquidity than dealer j ($V_i^s > V_j^s$), buyers with switching rates above some cutoff k^* prefer dealer i over dealer j, while buyers below the cutoff (those with $k < k^*$) prefer dealer j over dealer i. Thus, investors with greater liquidity needs (i.e. high k buyers) select dealers offering better liquidity, while investors with low liquidity needs select dealers offering low liquidity.

Lemma 1 (Sorting Device)**.** *Suppose k^* is such that $\hat{V}_i^b(k^*) = \hat{V}_j^b(k^*)$ for some dealers $i, j \in N$. Then, $\hat{V}_i^b(k) - \hat{V}_j^b(k)$ has the same sign as $(k - k^*)\left(V_i^s - V_j^s\right)$.*

Lemma 1 narrows down the set of equilibria into two main types based on the equilibrium heterogeneity in $\{V_i^s\}$: symmetric and asymmetric. We define an equilibrium as symmetric if all n dealers provide an identical liquidity service ($V_i^s = V_j^s$ for all i, j in N) and asymmetric if it features complete dealer heterogeneity ($V_i^s \neq V_j^s$ for all i, j in N). In the next two subsections, we discuss the existence of these equilibria, the distribution of clients across dealers, the resulting dealer heterogeneity, and the interdealer network structure.

3.1 Symmetric Equilibrium

Lemma 2 shows that a symmetric equilibrium exists. It has the following client and interdealer network structure. First, dealers do not specialize. Each dealer serves the entire spectrum of customers from the most buy-and-hold ($\underline{k}$) to the most liquidity investor ($\overline{k}$). This is because, using Lemma 1, investors are indifferent between the n homogenous dealers: $\nu_i(k) > 0$ for all $i \in N$ and all $k \in[\underline{k}, \overline{k}]$. Second, dealers have identical network centrality (M_i^{DD}). This is because dealers have identical client sizes. The left plot of Figure 5 illustrates the interdealer network structure. These results show that, first, we do not have any baked-in dealer heterogeneity. Second, client heterogeneity *alone* does not guarantee dealer heterogeneity.

Lemma 2 (Symmetric Equilibrium)**.** *A symmetric equilibrium exists in which dealers are homogenous in the liquidity service they provide to clients: $V_i^s = V_j^s$ for all $i, j \in N$. In such equilibria, dealers have identical client sizes and identical network centrality ($\mu_i^s = \mu_j^s$, $\mu_i^b = \mu_j^b$, $\mu_i^o = \mu_j^o$, $M_i^{DD} = M_j^{DD}$ for all $i, j \in N$).*

3.2 Main Results: Endogenous Specialization and Dealer Heterogeneity

This section gives the main results of the paper.

Theorem 1 shows that an asymmetric equilibrium exists in which no two dealers have identical liquidity service.[12] The heterogeneity in $\{V_i^s\}$ implies the following client structure across dealers. Let us index dealers in the order of increasing liquidity service: $V_1^s < V_2^s < \ldots < V_n^s$. This is without loss of generality. Then, according to Theorem 1, cutoffs $\{k_{1,2}^*, k_{2,3}^*, \ldots k_{n-1,n}^*\}$ exist where $\underline{k} < k_{1,2}^* < k_{2,3}^* < \ldots k_{n-1,n}^* < \overline{k}$, buyers of type $k < k_{1,2}^*$ choose dealer 1, buyers of type $k \in [k_{1,2}^*, k_{2,3}^*]$ choose dealer 2, and so on. Buyers at the cutoff $k = k_{i,j}^*$ are indifferent between dealers i and j: $V_i^b(k_{i,j}^*) = V_j^b(k_{i,j}^*)$. Thus, in the asymmetric equilibrium, dealers specialize. High liquidity quality dealers specialize in investors with frequent trading needs. Low liquidity dealers specialize in the relatively buy-and-hold investors. Figure 4 illustrates the clientele result.

Theorem 1 (Asymmetric Specialization Equilibrium)**.** *Suppose Assumption 10 holds. An asymmetric equilibrium exists in which $V_1^s < V_2^s < \ldots < V_n^s$. It is characterized by cutoffs $\{k_{1,2}^*, k_{2,3}^*, \ldots, k_{n-1,n}^*\}$, where $\underline{k} < k_{1,2}^* < \ldots k_{n-1,n}^* < \overline{k}$, buyers of type $k < k_{1,2}^*$ choose dealer 1, buyers of type $k \in [k_{1,2}^*, k_{2,3}^*]$ choose dealer 2, buyers of type $k \in [k_{2,3}^*, k_{3,4}^*]$ choose dealer 3 and so on. Buyers at the cutoff $k = k_{i,j}^*$ are indifferent between dealers i and j: $V_i^b(k_{i,j}^*) = V_j^b(k_{i,j}^*)$. In such equilibrium, $\mu_1^s < \mu_2^s < \ldots < \mu_n^s$ and $\mu_1^b < \mu_2^b < \ldots < \mu_n^b$.*

[12] In the Appendix, we show that the equilibrium is unique for $\lambda_{\rm D} = \lambda_{\rm DD}$. Showing uniqueness for more general parameter values is tedius.

Dealers are heterogeneous not only in the composition of clients but also in the size of client masses or, equivalently, in the volume of client orders they receive. From Theorem 1, high liquidity dealers receive larger volumes of client orders, while low liquidity dealers receive smaller volumes.

The heterogeneity in the volume of client orders generates dealer heterogeneity on the interdealer market. Theorem 2 shows that dealers specializing in liquidity investors—supported by their large volume of client orders—intermediate larger volumes of interdealer trades and consequently form the core. The mechanism reverses for peripheral dealers. The intuition is simple. If a dealer receives a large volume of client orders, she intermediates a large number of bonds both by herself (i.e. in-house) and by trading with other dealers (i.e. inter-house). Thus, how clients form around dealers determines the shape of the interdealer network. It determines the volume of orders dealers receive and, as a result, how much dealers trade on the interdealer market. This is the main insight of the paper. Figure 5 illustrates the network structure in the asymmetric equilibrium.

Theorem 2 (An Endogenous Core-Periphery Network)**.** *The dealers that specialize in investors with relatively high switching rates (i.e. high liquidity need investors) intermediate more CDC chains:* $M_n^D > M_{n-1}^D > \ldots > M_2^D > M_1^D$*. They also intermediate more interdealer (i.e. CDDC) trades,* $M_n^{DD} > M_{n-1}^{DD} > \ldots > M_2^{DD} > M_1^{DD}$*, and, as a result, form the core of the interdealer network. It is vice versa for dealers that specialize in investors with relatively low switching rates (i.e. buy-and-hold investors).*

3.3 Intuition

This section provides an intuition for (1) the tradeoffs clients face in choosing between dealers; (2) why high liquidity dealers are also the dealers with larger client masses; and (3) the mechanism generating the heterogeneity in $\{V_i^s\}$.

For the clientele effect to arise, choosing the high liquidity dealer has to come at a cost. Proposition 1 highlights the cost and thereby the tradeoff clients face in choosing between dealers.

Proposition 1 (Properties of the Asymmetric Equilibrium)**.** *Suppose dealers* i *and* j *are such that dealer* i *provides better liquidity:* $V_i^s > V_j^s$*. If* $\lambda_{\text{D}} \geq \lambda_{\text{DD}}$ *and* r *is small, dealer* i *charges a higher expected ask-price than dealer* j*:* $\overline{p}_i^{ask}(k) > \overline{p}_j^{ask}(k)$ *for all* $k \in [\underline{k}, \overline{k}]$ *.*

Investors tradeoff a dealer's liquidity service, V_i^s, versus the price they expect to pay on average for the bond, $\overline{p}_i^{ask}(k)$. Some dealers charge a high ask-price on average but, in return, offer better liquidity if the client has to return the bond: the dealer either buys back at a higher bid-price, executes orders more quickly, or both. Others charge a cheaper ask-price but offer worse liquidity. Buyers who expect to reverse their position quickly (i.e., those with high switching rates, k) care more about what happens to them as a seller. They, as a result, choose the dealer based on its liquidity 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. Figure 6 illustrates the tradeoff. Appendix C.6 explains the tradeoff in detail.

The heterogeneity in client sizes across dealers generates these tradeoffs. The tradeoffs arise, in particular, if the high liquidity dealer is the larger client mass dealer. To see the intuition, consider for simplicity the parameter range where dealers have identical execution speed: $\lambda_{\text{DD}} = \lambda_{\text{D}}$. In this case, if the high liquidity dealer has a larger client mass, she charges higher expected ask-prices. The mechanism has two parts. First, the ask-price a large client mass dealer charges in in-house matches dictates its average ask-price. This is because the larger client mass dealer fills a larger fraction of its client orders by itself (i.e. in-house) than by involving another dealer (inter-house). Second, when $\lambda_{\text{DD}} = \lambda_{\text{D}}$, in-house matches result in higher (hence, worse) prices for buyers than inter-house matches regardless of the dealer. This is because when $\lambda_{\text{DD}} \leq \lambda_{\text{D}}$, Assumption 1 requires $z_{\text{DD}} > z_{\text{D}}$, which says that clients have less bargaining power (and thereby extract smaller gains from trade) in one-dealer chains than in two-dealer chains. A buyer, as a result, pays a higher price (a price closer to her reservation value) in one-dealer chains than in two-dealer chains. Now, recall that the dealer with a larger client size is the relatively core dealer on the interdealer market. Put together, if an investor becomes a client of a core dealer, she buys on average through an in-house match, has less bargaining power against her dealer, and thereby pays a higher ask-price. The opposite holds for a peripheral dealer. Their clients buy on average through an inter-house match, have more leverage against their dealer (due to $z_{\text{DD}} > z_{\text{D}}$), and thereby buy at a cheaper ask-price. The core dealer, as a result, is the higher cost dealer. But since the higher cost has to come with a larger benefit, the core dealer must be the higher liquidity dealer. Thus, the heterogeneity in client sizes ensures that choosing the larger client mass dealer involves both a larger benefit and a larger cost.

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 continuation value 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 (10), 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.[13] 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.

3.4 Key Ingredients

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

The second ingredient is the parameter condition in (10): $\lambda_{\text{DD}} z_{\text{DD}} > \lambda_{\text{D}} z_{\text{D}}$. 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 micro-foundations for why intermediation chains are beneficial. We, instead, capture them in a reduced form through (10). Glode and

[13] 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.

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 micro-foundations that generate the parameter conditions.[14]

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.[15]

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.

4 Testable Predictions

We now tie the network centrality results with the previous results on specialization. We highlight testable predictions of our model and compare them with the available empirical evidence.

4.1 Client Trades

Our model has a broader interpretation. 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

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

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

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.[16]

4.2 CDC and CDDC Chains

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

$$M_c^D > M_p^D \tag{13}$$

Thus, core dealers account for a larger fraction of not only interdealer trades (hence their labels) but also client trades.[17] 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.[18] 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 are precisely why they form the core.

A core dealer intermediates more CDC chains both in levels as in (13) and as a fraction of all chains it intermediates*:*

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

Thus, a core dealer intermediates client trades more on its own than by relying on the inter-dealer market. A peripheral dealer, in contrast, relies more on other dealers and hence on long intermediation chains for its liquidity service to clients.[19] For a peripheral dealer, CDDC chains

[16]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.

[17]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.

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

[19]To see this, multiply (13) by negative 1 and add 1 to both sides.

comprise a larger fraction of all its intermediations than for a core dealer:

$$\frac{M_p^{DD}}{M_p^D + M_p^{DD}} > \frac{M_c^{DD}}{M_c^D + M_c^{DD}}. \tag{15}$$

Eq. (14) also implies that the average chain length is longer for a peripheral dealer:

$$\frac{M_p^D}{M_p^D + M_p^{DD}}(1) + \frac{M_p^{DD}}{M_p^D + M_p^{DD}}(2) > \frac{M_c^D}{M_c^D + M_c^{DD}}(1) + \frac{M_c^{DD}}{M_c^D + M_c^{DD}}(2), \tag{16}$$

where inside the brackets are the chain lengths. LS and NHS document the same patterns as (13)-(16).

4.3 Execution Speed

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

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

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.[20] The rate of filling sell orders is analogously defined as:

$$m_i^b \equiv \frac{M_i^D + 0.5M_i^{DD}}{\mu_i^s} = \sum_{j\in N} \lambda_{ij}\mu_j^b \tag{18}$$

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

$$\begin{aligned} & m_i^\tau - m_j^\tau = \\ & = \left[\lambda_{\text{D}}\mu_i^\tau + \lambda_{\text{DD}}\mu_j^\tau + \lambda_{\text{DD}} \sum_{j\in N/\{i,j\}} \mu_j^\tau\right] - \left[\lambda_{\text{D}}\mu_j^\tau + \lambda_{\text{DD}}\mu_i^\tau + \lambda_{\text{DD}} \sum_{j\in N/\{i,j\}} \mu_j^\tau\right] \\ & = -(\lambda_{\text{DD}} - \lambda_{\text{D}})\left(\mu_i^\tau - \mu_j^\tau\right) \end{aligned} \tag{19}$$

for $\tau = \{s, b\}$. Thus, if $\lambda_{\text{DD}} < \lambda_{\text{D}}$, a core dealer executes at a faster rate: $m_c^\tau > m_p^\tau$ for $\tau = \{s, b\}$. If $\lambda_{\text{DD}} = \lambda_{\text{D}}$, a core dealer offers the same execution speed as a peripheral dealer: $m_i^\tau = m_j^\tau$. If $\lambda_{\text{DD}} > \lambda_{\text{D}}$, a core dealer executes at a slower rate: $m_c^\tau < m_p^\tau$. The intuition for the latter

[20] 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 $m_i^s dt$.

is: A core dealer fills large volumes of client orders (the numerator in (17) and (18)), 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.[21]

4.4 Transacted and Quoted Ask- and Bid-Prices

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

$$E_c^b\left[\overline{p}_c^{ask}(k)\right] < E_p^b\left[\overline{p}_p^{ask}(k)\right] \tag{20}$$

Eq. (20) 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:

$$\overline{p}_c^{ask}(k) > \overline{p}_p^{ask}(k) \tag{21}$$

Eq. (21) 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 (20), not (21). Thus, (20) is the testable prediction relevant to transaction price data, not (21).[22]

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.

[21] 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.

[22] In a data with both quoted and transaction prices, it is possible to check (21) also.

4.5 Transacted Bid-Ask Spread

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

$$E_c^b\left[\overline{\phi}_c(k)\right] < E_p^b\left[\overline{\phi}_p(k)\right], \tag{22}$$

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. (22) can be seen in Figure 7. 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 (22).

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.

4.6 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 7 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 sufficiently high λ_{DD}, the bid-price is so high that the core dealer offers a narrower bid-ask spread for all k.[23] On the other extreme, 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 $k \in (\tilde{k}, \overline{k}]$.[24]

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

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

4.7 Interdealer Trades

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

Proposition 2 (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.[25] 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:

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

It is analogous for dealer d's buy-side trades.[26]

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.[27] The end-seller's high reservation value,

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

[26]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.

[27]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.

in turn, manifests as a high interdealer ask-price.[28] 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).[29]

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.

5 Additional Results

In Sections 5.1 and 5.2, we analyze how dealer interconnectedness and market fragmentation affects prices, liquidity, and welfare. In Section 5.3, we analyze the welfare across asymmetric and symmetric equilibria.

We start by characterizing customer welfare, dealer profits, and the total welfare. We define customers' welfare as

$$W^C \equiv \sum_{i \in N} [\int_{\underline{k}}^{\overline{k}} \hat{\mu}_i^b(k) V_i^b(k) dk + \int_{\underline{k}}^{\overline{k}} \hat{\mu}_i^o(k) V_i^o(k) dk + \mu_i^s V_i^s. \tag{23}$$
$$+ \frac{1}{r} \int_{\underline{k}}^{\overline{k}} V_i^b(k) \hat{f}(k) \nu_i(k) dk]$$

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

[29]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.

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

$$\begin{aligned} W_i^D \equiv &\frac{1}{r}\int_{\underline{k}}^{\overline{k}} \lambda_{\mathrm{D}}\hat{\mu}_i^b(k)\mu_i^s\left(1-2z_{\mathrm{D}}\right)\left(V_i^o(k)-V_i^b(k)-V_i^s\right)dk \\ &+\frac{1}{r}\sum_{j\in N_i}\left(\int_{\underline{k}}^{\overline{k}} \lambda_{\mathrm{DD}}\hat{\mu}_i^b(k)\mu_j^s\left(\frac{1-2z_{\mathrm{DD}}}{2}\right)\left(V_i^o(k)-V_i^b(k)-V_j^s\right)dk\right) \\ &+\frac{1}{r}\sum_{j\in N_i}\left(\int_{\underline{k}}^{\overline{k}} \lambda_{\mathrm{DD}}\hat{\mu}_j^b(k)\mu_i^s\left(\frac{1-2z_{\mathrm{DD}}}{2}\right)\left(V_j^o(k)-V_j^b(k)-V_i^s\right)dk\right). \end{aligned} \tag{24}$$

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

$$W^D \equiv \sum_{i\in N} W_i^D. \tag{25}$$

The total welfare of all agents in the economy is then

$$W_{all} \equiv W^C + W^D. \tag{26}$$

As Proposition 3 shows, the total welfare depends only on the aggregate mass of sellers, $\mu_N^s = \sum_{i\in N}\mu_i^s$.

Proposition 3. *The total welfare is given by*

$$W_{all} = \frac{\delta}{r}S - \frac{x}{r}\mu_N^s. \tag{27}$$

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 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.

To compare prices and bid-ask spreads across different environments, we take the weighted average across dealers:

$$\bar{p}_{ask} \equiv \frac{1}{\sum_{i\in N}\left(\frac{1}{2}M_i^{DD}+M_i^D\right)}\sum_{i\in N}\left[\left(\frac{1}{2}M_i^{DD}+M_i^D\right)E_i^b\left[\overline{p}_i^{ask}(k)\right]\right] \tag{28}$$

$$\bar{p}_{bid} \equiv \frac{1}{\sum\limits_{i\in N}\left(\frac{1}{2}M_i^{DD}+M_i^D\right)}\sum_{i\in N}\left[\left(\frac{1}{2}M_i^{DD}+M_i^D\right)\overline{p}_i^{bid}\right] \tag{29}$$

$$\bar{\phi} \equiv \frac{1}{\sum\limits_{i\in N}\left(\frac{1}{2}M_i^{DD}+M_i^D\right)}\sum_{i\in N}\left[\left(\frac{1}{2}M_i^{DD}+M_i^D\right)E_i^b\left[\overline{\phi}_i(k)\right]\right]. \tag{30}$$

5.1 Dealer Interconnectedness

In this section, we contrast the environments with and without the interdealer market and show that dealer interconnectedness through the interdealer market increases bond market liquidity and the total welfare. Without the interdealer market, clients trade only locally with the clients of the same dealer. This environment is similar to Vayanos and Wang (2007).[30] Markets in their setting are the counterparts to dealers in our setting. We assume the supply of bonds circulating among customers of each dealer is identical at $s_i = S/n$.

In the absence of interdealer trades, clients sort into dealers analogous to the environment with interdealer trades. Buyers tradeoff a dealer's expected ask-price versus its liquidity service. Buy-and-hold investors choose the dealer offering a cheaper price, while liquidity investors choose the dealer offering better liquidity. Dealers specializing in high liquidity need investors have a larger buyer and seller client mass than dealers specializing in low liquidity need investors.

Proposition 4 (The Effect of Interconnectedness)**.** *Dealer interconnectedness increases bond market liquidity (the aggregate volume of trade and execution speeds), improves the allocation of bonds (μ_N^s), and increases the total welfare (W_{all}).*

The intuition for Proposition 4 is as follows. Due to interdealer trades, an investor trades not only with the other clients of her dealer but also with the entire investor population in the economy. Each dealer, as a result, intermediates between much larger pools of buyers and sellers, produces greater volumes of trade, and fills client orders at a higher speed. The increased market liquidity improves the allocation of bonds. Bond owners upon reverting to a low-valuation investor now sell more quickly to another high-valuation investor. A larger number of high-valuation investors, as a result, hold the bond. The improved asset allocation increases the total welfare.[31]

[30] The model in Vayanos and Wang (2007) is a special case of our model with $z_{ij} = 1$ and $N_i = \{\emptyset\}$ for all i.

[31] As Appendix G.1 discusses, how interconnectedness affects clients' welfare, dealer profits, and prices depends on clients' bargaining power.

As Proposition 5 shows, dealer interconnectedness evens out the order imbalance across dealers. We capture the order imbalance of a dealer by the ratio of its seller mass to its buyer mass. Without the interdealer market, the ratio differs across dealers and is higher for dealers specializing in buy-and-hold investors. Introducing the interdealer market removes the heterogeneity. Imbalances at the dealer level are now proportional to the aggregate imbalance. Thus, dealers achieve a full risk-sharing in their orders.

Proposition 5. *In the presence of the interdealer market, the order balance is identical across dealers: for all $i \in N$,*

$$\frac{\mu_i^s}{\mu_i^b} = \frac{\mu_N^s}{\mu_N^b}. \tag{31}$$

5.2 Market Fragmentation

In this section, we analyze how market fragmentation affects customer welfare, dealer profits, prices, and liquidity. We capture market fragmentation with the aggregate number of dealers in the economy (n) and, as in the main section, assume a complete network. We focus on the parameter range $\lambda_{\text{DD}} > \lambda_{\text{D}}$, which is the more interesting range, and relegate to the appendix the results under $\lambda_{\text{DD}} \leq \lambda_{\text{D}}$.

Proposition 6. *Suppose $\lambda_{\text{DD}} > \lambda_{\text{D}}$. Then, increasing the number of dealers in the economy increases the aggregate volume of trade, $\sum_{i \in N} \left(M_i^D + M_i^{DD}\right)$ and the total welfare in the economy, W_{all}.*

The intuition for Proposition 6 is as follows. With more dealers in the economy, each dealer has a smaller client mass and, as a result, is more likely to involve another dealer to fill a client's order. But due to condition $\lambda_{\text{DD}} > \lambda_{\text{D}}$, two dealers working together produce more matches than a dealer working alone. The result is a greater volume of trade, a better allocation of bonds, and a higher social welfare.

As Figure 10 illustrates, how clients and dealers split the increase in the welfare depends on clients' bargaining power in two-dealer chains (z_{DD}) relative to that in one-dealer chains (z_{D}). If it is significantly larger in two-dealer chains, clients not only extract the entire increase in the welfare but also get a cut from dealers' profit. That is, by lengthening the intermediation chain, clients tilt the gains from trade in their favor at the expense of dealers. Dealers in this case are better off in a concentrated market with as few other dealers as possible. For an intermediate range of clients' bargaining power in two-dealer chains, both clients and dealers benefit from fragmentation. Finally, if clients' bargaining power is smaller in two-dealer chains than in one-dealer chains, now dealers extract the increase in the welfare. They do so at the expense of customer welfare.

These changes in how clients and dealers split the gains from trade correspond to changes in the bid-ask spreads clients pay. In particular, a decrease in dealer profits manifests as a decrease in the bid-ask spreads clients face and vice versa for an increase in dealer profits.

5.3 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 concentration. For exposition, we do so in an environment with two dealers, where dealer 1 serves the low switching rate buyers, and dealer 2 serves the high switching rate buyers.

Figure 12 illustrates how the level of heterogeneity (captured by $\frac{\mu_2^s}{\mu_1^s}$) changes with the cutoff. Let us denote by k^*_{sym} a cutoff such that the two dealers are identical ($\mu_1^s = \mu_2^s$). Even though this is not an equilibrium cutoff, the client masses at this cutoff correspond to the client masses of the symmetric equilibrium.[32] Then, for any cutoff $k^* < k^*_{sym}$, dealer 2 is the larger dealer. In particular, the lower the cutoff is, the larger is the heterogeneity between the dealers ($\frac{\mu_2^s}{\mu_1^s}$). This is because dealer 1's client mass increases with the cutoff, while dealer 2's decreases. Moreover, the asymmetric equilibrium cutoff, k^*_{asym}, has to be below k^*_{sym}. This is because the dealer that specializes in high switching rate buyers (i.e. dealer 2) has to be the larger dealer in the asymmetric equilibrium.

Proposition 7. *Suppose $\lambda_{\text{DD}} > \lambda_{\text{D}}$. Then, the socially optimal cutoff (k^*_{wel}) is such that dealers are heterogeneous: $\mu_1^s < \mu_2^s$.*

According to Proposition 7, 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, 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

[32]This is not an equilibrium cutoff because for dealers to specialize in equilibrium, dealers necessarily have to be heterogenous.

precisely that. A dealer specializing in buy-and-hold investors provides a faster liquidity immediacy than a dealer specializing in liquidity investors. Importantly, Proposition 7 implies that dealer specialization and a core-periphery network are socially desirable. Figure 12 illustrates with k^*_{wel} the cutoff that maximizes the total welfare W_{all}.

Next, consider how the equilibrium level of dealer heterogeneity compares with the socially optimal level. Let us allow buyers and sellers' bargaining powers to differ because they affect the equilibrium heterogeneity in opposite directions. Let us also focus on the parameter range $z_{\text{DD}} > z_{\text{D}}$. As Figure 12 illustrates, increasing the buyers' bargaining power in inter-house matches (holding the other parameters fixed) reduces the equilibrium dealer heterogeneity. The intuition is as follows. If we increase buyers' bargaining power in inter-house matches (without changing client masses), the marginal cost of choosing the larger dealer increases, while the marginal benefit remains the same. Fewer buyers, as a result, choose the larger dealer (dealer 2). As dealer 2's client size decreases (while dealer 1's size increases), dealers become less heterogenous. Reducing the dealer heterogeneity, in turn, starts to reverse the increase in the marginal cost ($\overline{p}^{ask}_2(k) - \overline{p}^{ask}_1(k)$) while also increasing the marginal benefit. This process continues until there is a buyer who is indifferent between the two dealers. The result is more homogenous dealers. If buyers' bargaining power is sufficiently large, dealer heterogeneity (thereby, the extent of the core-periphery structure and concentration on the interdealer market) falls short of the socially optimal level.

Sellers' bargaining power has the opposite effect (as explained further in the Appendix). Dealer heterogeneity increases with the sellers' bargaining power. If sellers' bargaining power is sufficiently large, the equilibrium dealer heterogeneity (hence, the extent of concentration on the inter-dealer market) can exceed the socially optimal level.

The discussion so far focuses on the parameter range $\lambda_{\text{DD}} > \lambda_{\text{D}}$. If instead $\lambda_{\text{DD}} < \lambda_{\text{D}}$, then pooling all clients under one dealer (that is, merging dealers into a single monopoly dealer) maximizes the social welfare. If $\lambda_{\text{DD}} = \lambda_{\text{D}}$, the aggregate mass of sellers and thereby the total social welfare does not depend on how clients are organized around dealers.[33] That is, the symmetric and asymmetric equilibria yield the same welfare, and a social planner cannot improve on them.

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

6 Discussion

In this section, we discuss our assumptions and how relaxing them would affect our results. In Section 3, 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.

7 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.

Appendix

A 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]
Supply of bonds	S	0.3
Risk-free rate	r	0.04

B Figures

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. See Section 2 for more detail.

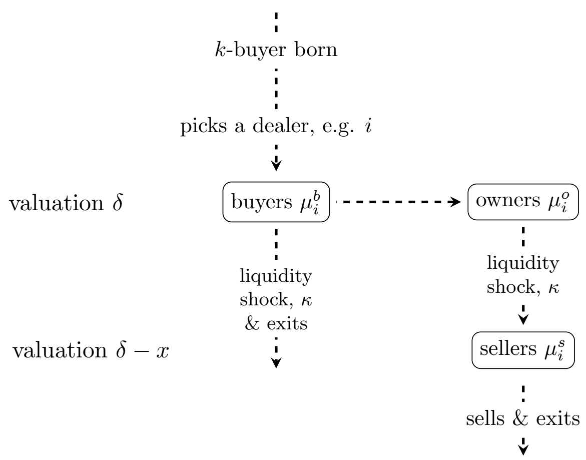


Figure 2: Clients, Dealers, and Interdealer Trades

The figure illustrates the model environment, as an example, for $n = 3$ dealers. 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. See Section 2 for more detail.

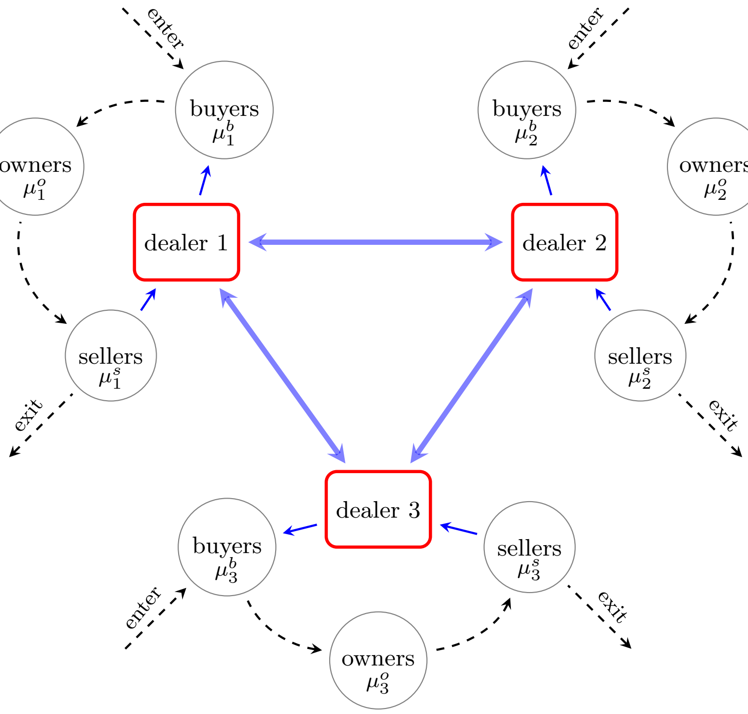


Figure 3: Transaction Prices

The figure illustrates how prices arise from a sharing rule. The total gains from trade is the difference between the end-buyer's reservation value $(V_j^o(k) - V_j^b(k))$ and end-seller's reservation value (V_i^s). Prices, characterized in (C.7)-(C.9), are such that the two end-customers each capture z_{D} (z_{DD}) fraction of the total surplus in one-dealer (two-dealer) intermediation chains; dealer(s) split equally the remaining $1 - 2z_{\text{D}}$ $(1 - 2z_{\text{DD}})$ fraction. The top plot illustrates prices in one-dealer intermediation chains, while the bottom plot illustrates prices for two-dealer intermediation chains. See Sections 2.6 and C.3 for more detail.

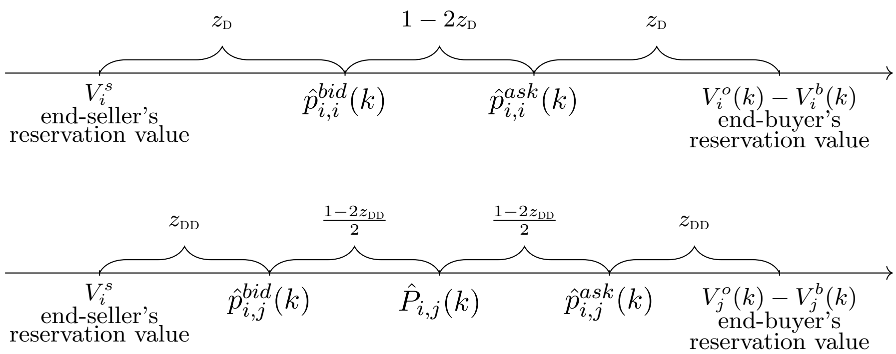


Figure 4: Endogenous Cutoffs and Specialization in Customers

The figure illustrates how in the asymmetric equilibrium dealers specialize in different clients with different liquidity needs (k). See Section 3.2 for more detail.

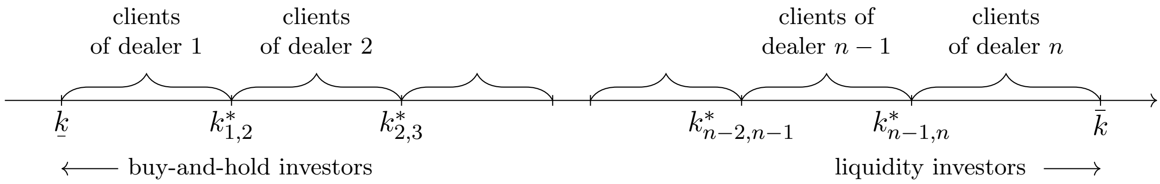


Figure 5: An Endogenous Core-Periphery Structure

The figure illustrates the equilibrium interdealer network structure in the symmetric (the left plot) and in the asymmetric equilibria (the right plot). Each node is a dealer. The sizes of nodes represent the buyer and seller-client sizes of dealers. Lines represent trades between dealers. The thickness of the lines represent per period volume of trade between pairs of dealers. The symmetric equilibrium features dealers who are identical in terms of their client size and network centrality. The asymmetric equilibrium exhibits a core-periphery interdealer network. See Sections 3.1-3.2 for more detail.

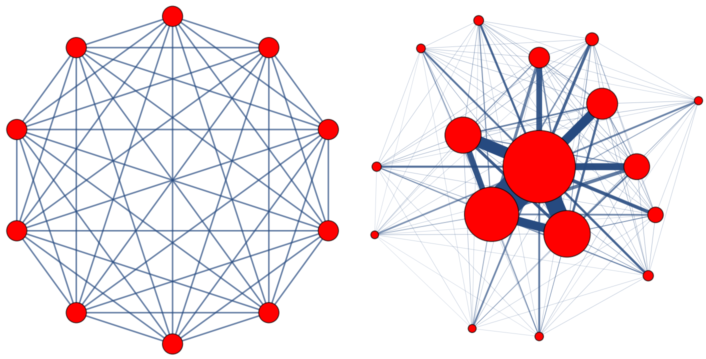


Figure 6: The Average Ask-Price vs. the Sellers' Expected Utility

The figures illustrate the tradeoff that buyers face when choosing dealers, for exposition, in a two-tier environment. The figures plot the expected ask-price (on the left) and the continuation value of a seller-client (on the right) as functions of clients' liquidity type k (in x-axis) for a relatively core versus central dealer. The expectation of the ask-price is over possible sellers that a buyer could be matched with as a client of that dealer. The cutoff k^* is such that buyers with $k < k^*$ choose the peripheral dealer, while buyers with $k \geq k^*$ choose the core dealer. See Section 3.3 for more detail.

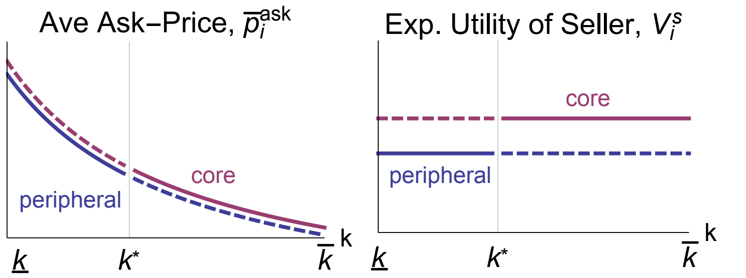


Figure 7: The Quoted Average Bid-Ask Spread

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: $\lambda_{\text{DD}} < \lambda_{\text{D}}$, $\lambda_{\text{DD}} = \lambda_{\text{D}}$, $\lambda_{\text{DD}} > \lambda_{\text{D}}$. The cutoff k^* is such that buyers with $k < k^*$ choose the peripheral dealer, while buyers with $k \geq k^*$ choose the core dealer. See Sections 4.5-4.6 for more detail.

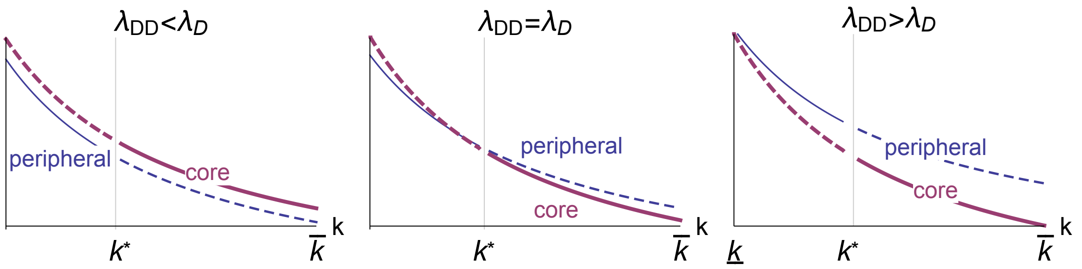


Figure 8: The Effect of Dealer Interconnectedness on Welfare

The figures illustrate how the introduction of the interdealer market changes the total welfare, dealer profits, and clients' welfare. They illustrate the change in these variables across the symmetric equilibria and as functions of clients' bargaining power (z_{DD}) and matching efficiency (λ_{DD}) in two-dealer chains. The parameter values used for the plot are $z_{\text{D}} = 0.1$ and $z_{\text{D}} = 0.3$ in the environments with and without interdealer trades, respectively, and $\lambda_{\text{D}} = 100$ for both. The other parameter values are in Table 1. See Section 5.1 for more detail.

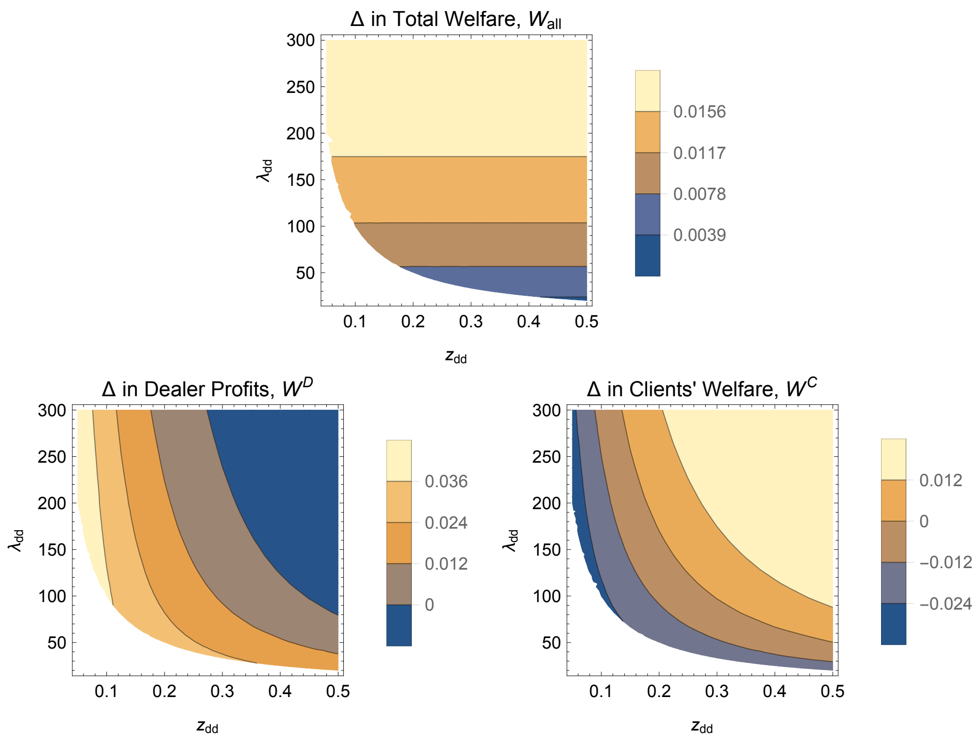


Figure 9: The Effect of Dealer Interconnectedness on Prices

The figures illustrate how the introduction of the interdealer market changes the average bid-ask spread, the average ask-price, and the average bid-price. They illustrate the change in these variables across the symmetric equilibria and as functions of clients' bargaining power (z_{DD}) and matching efficiency (λ_{DD}) in two-dealer chains. The parameter values used for the plot are $z_{\text{D}} = 0.1$ and $z_{\text{D}} = 0.3$ in the environments with and without interdealer trades, respectively, and $\lambda_{\text{D}} = 100$ for both. The other parameter values are in Table 1. See Section 5.1 for more detail.

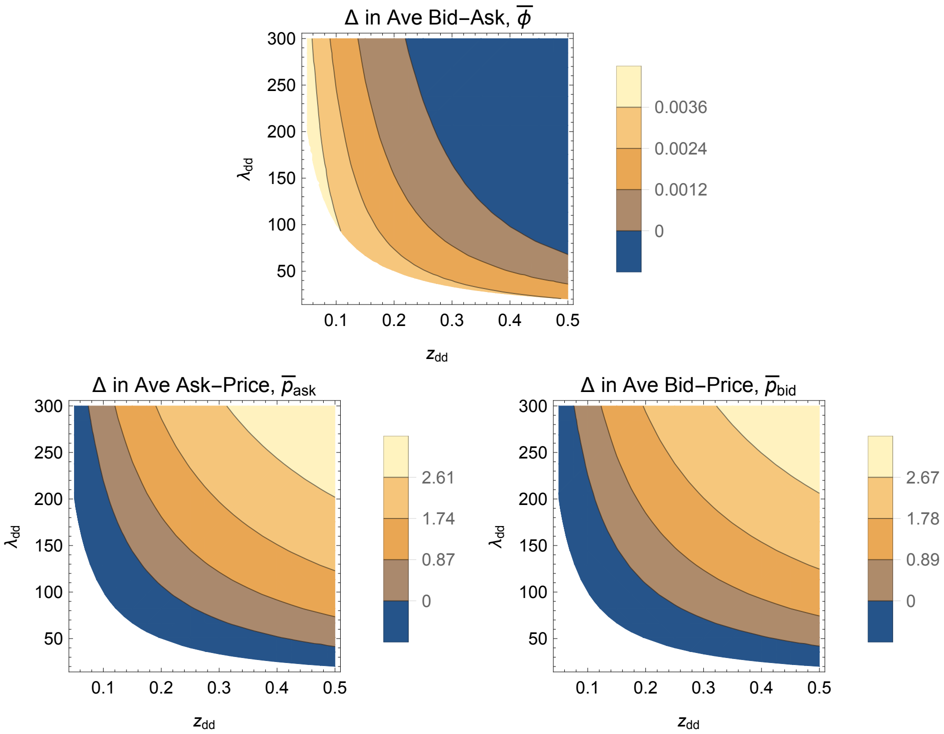


Figure 10: The Effect of Market Fragmentation on Welfare

The figures illustrate the change in the total welfare, dealer profits, and clients' welfare as the number of dealers increase from $n = 2$ to $n = 3$. They illustrate the change in these variables across symmetric equilibria and as functions of clients' bargaining power (z_{DD}) and matching efficiency (λ_{DD}) in two-dealer chains. The parameter values used for the plot are $z_D = 0.2$ and $\lambda_D = 100$. The other parameter values are in Table 1. See Section 5.2 for more detail.

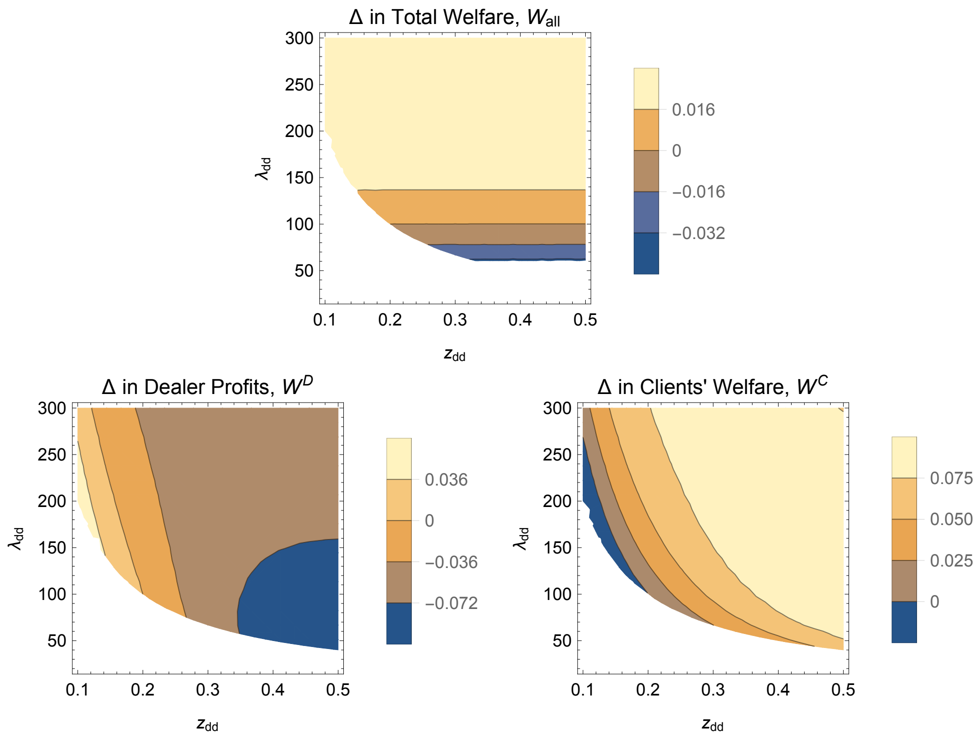


Figure 11: The Effect of Market Fragmentation on Prices

The figures illustrate the change in the average bid-ask spread, the average ask-price, and the average bid-price as the number of dealers increase from $n = 2$ to $n = 3$. They illustrate the change in these variables across symmetric equilibria and as functions of clients' bargaining power (z_{DD}) and matching efficiency (λ_{DD}) in two-dealer chains. The parameter values used for the plot are $z_D = 0.2$ and $\lambda_D = 100$. The other parameter values are in Table 1. See Section 5.2 for more detail.

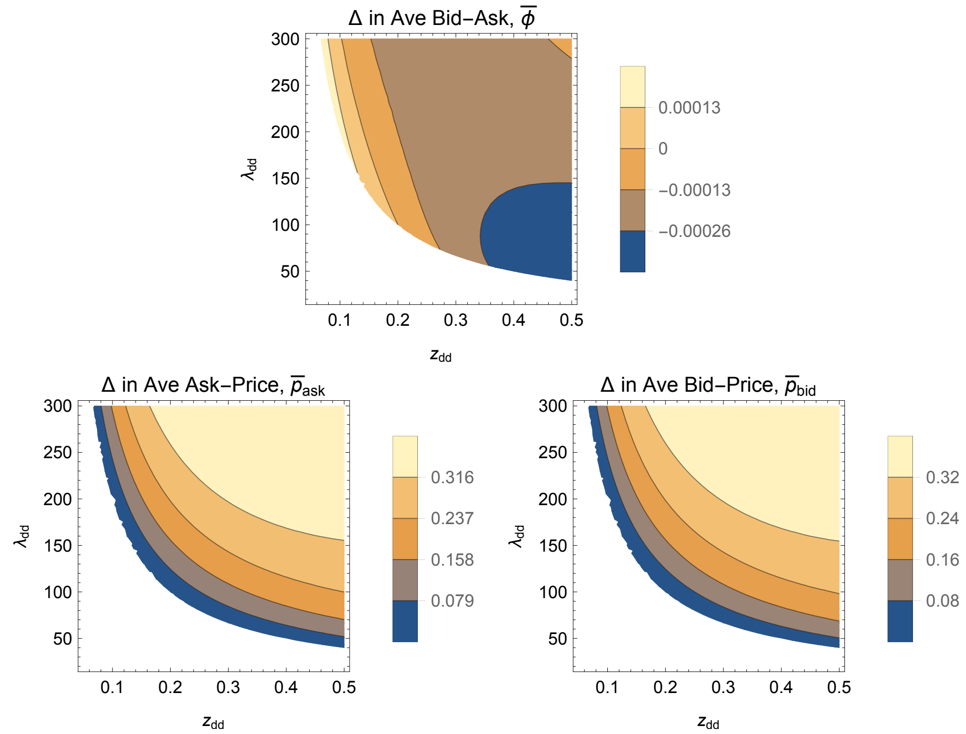
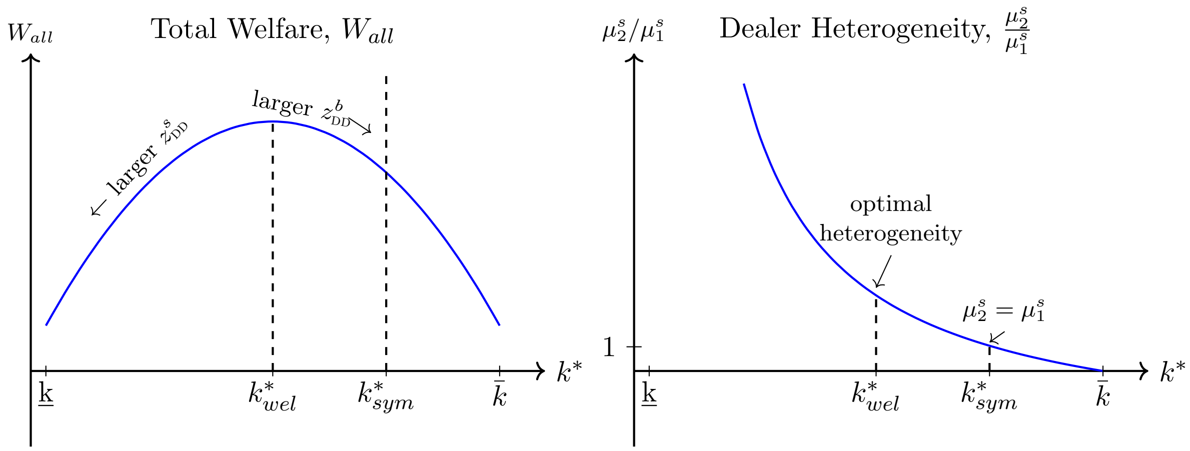


Figure 12: Welfare Analysis

The figure plots the total welfare (on the left) and the level of dealer heterogeneity (on the right) as functions of the cutoff k^*. It does so for a two-tier environment in which $\lambda_{DD} > \lambda_D$. The cutoff k^*_{sym} is a cutoff such that $\mu^s_1 = \mu^s_2$, and k^*_{wel} is a cutoff that maximizes the total welfare. The actual asymmetric equilibrium cutoff, k^*_{asym}, will be in $(\underline{k}, k^*_{sym})$. Where it falls depends on clients' bargaining power. If the sellers' bargaining power in two-dealer chains (z^s_{DD}) is large, the cutoff is smaller implying more heterogenous dealers. If the buyer's bargaining power in two-dealer chains (z^b_{DD}) is large, the cutoff is larger (implying more homogenous dealers) and can even exceed k^*_{wel}. See Section 5.3 for more detail.



C Masses, Value Functions, Prices, Liquidity, and Tradeoffs

C.1 Client Masses

The mass of k-type owners is given by

$$\left(\sum_{j\in N}\lambda_{ij}\mu_j^s\right)\hat{\mu}_i^b(k) = k\hat{\mu}_i^o(k). \tag{C.1}$$

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.

Later in the proofs, we will use the characterization of client masses given in the following Lemma.

Lemma C.1. *Taking the cutoff functions $\{\nu_i(k)\}_i$ as given, the system of equations characterizing $\{\mu_i^s\}_i$ is given by*

$$\int_{\underline{k}}^{\overline{k}} \frac{f(k)}{k + (\lambda_{\text{D}} - \lambda_{\text{DD}})\mu_i^s + \lambda_{\text{DD}}\mu_N^s}\nu_i(k)dk = \frac{\mu_i^s\left(c + \mu_N^s\right)}{\mu_N^s}$$

for each dealer i, where $\mu_N^s = \sum_{i\in N}\mu_i^s$ and $c \equiv \int_{\underline{k}}^{\overline{k}} f(k)\frac{1}{k}dk - S > 0$ is a constant.

Proof. The market clearing condition is

$$\sum_{i\in N} m_i^s\mu_i^o + \sum_{i\in N}\mu_i^s = S, \tag{C.2}$$

where

$$\mu_i^o = \int_{\underline{k}}^{\overline{k}} \frac{1}{k}\hat{\mu}_i^b(k)dk$$

and

$$m_i^s = (\lambda_{\text{D}} - \lambda_{\text{DD}})\mu_i^s + \lambda_{\text{DD}}\mu_N^s.$$

We can write

$$\begin{aligned}
\mu_i^b &= \int_{\underline{k}}^{\overline{k}} \frac{f(k)}{k + m_i^s}\nu_i(k)dk \\
&= \int_{\underline{k}}^{\overline{k}} \frac{1}{k}\frac{k + m_i^s - m_i^s}{k + m_i^s}\nu_i(k)f(k)dk \\
&= \int_{\underline{k}}^{\overline{k}} \frac{f(k)}{k}\nu_i(k)dk - \int_{\underline{k}}^{\overline{k}} \frac{1}{k}\frac{m_i^s}{k + m_i^s}\nu_i(k)f(k)dk \\
&= \int_{\underline{k}}^{\overline{k}} \frac{1}{k}\nu_i(k)f(k)dk - m_i^s\mu_i^o
\end{aligned}$$

From this,

$$m_i^s\mu_i^o = \int_{\underline{k}}^{\overline{k}} \frac{1}{k}\nu_i(k)f(k)dk - \mu_i^b$$

Substituting this back into the market clearing condition, we get:

$$\int_{\underline{k}}^{\overline{k}} \frac{1}{k}f(k)dk - \mu_N^b + \mu_N^s = S. \tag{C.3}$$

Eq. (C.3) together with system of interdealer constraints, $\mu_i^s = \mu_i^b \frac{\mu_N^s}{\mu_N^b}$ for all $i \in N$, characterize $\{\mu_i^s\}_i$ taking as given the dealer choice functions $\{\nu_i(k)\}_i$. □

C.2 Value Functions

After simplifying and taking the continuous time limit of (11), we get

$$rV_i^b(k) = k\left(0 - V_i^b(k)\right) + \sum_{j\in N} \lambda_{ij}\mu_j^s \left(V_i^o(k) - V_i^b(k) - \hat{p}_{j,i}^{ask}(k)\right). \tag{C.4}$$

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

$$rV_i^o(k) = \delta + k\left(V_i^s - V_i^o(k)\right). \tag{C.5}$$

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

$$rV_i^s = \delta - x + \sum_{j\in N}\left(\int_{\underline{k}}^{\overline{k}} \lambda_{ij}\hat{\mu}_j^b(k)(\hat{p}_{i,j}^{bid}(k) - V_i^s)dk\right). \tag{C.6}$$

C.3 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

$$\hat{p}_{i,j}^{bid}(k) = (1 - z_{ij})V_i^s + z_{ij}(V_j^o(k) - V_j^b(k)) \tag{C.7}$$

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:

$$\hat{P}_{i,j}(k) = \frac{1}{2}V_i^s + \frac{1}{2}(V_j^o(k) - V_j^b(k)). \tag{C.8}$$

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

$$\hat{p}_{i,j}^{ask}(k) = z_{ij}V_i^s + (1 - z_{ij})\left(V_j^o(k) - V_j^b(k)\right). \tag{C.9}$$

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.

C.4 Expected Prices and Liquidity from 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:

$$\overline{p}_i^{ask}(k) \equiv \frac{1}{m_i^s}\sum_{j\in N} \lambda_{ij}\mu_j^s \hat{p}_{j,i}^{ask}(k) \tag{C.10}$$

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:

$$p_i^{ask} \equiv E_i^b \left[\overline{p}_i^{ask}(k) \right], \tag{C.11}$$

where the expectation is over the buyer population measure.[34]

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 both dealer i and dealer i's connections):

$$\overline{p}_i^{bid} \equiv \frac{1}{m_i^b} \sum_{j\in N} \lambda_{ij}\mu_j^b E_j^b[\hat{p}_{i,j}^{bid}(k)], \tag{C.12}$$

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:

$$\overline{\phi}_i(k) \equiv \frac{\overline{p}_i^{ask}(k) - \overline{p}_i^{bid}}{0.5(\overline{p}_i^{ask}(k) + \overline{p}_i^{bid})}. \tag{C.13}$$

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

$$\phi_i \equiv \frac{p_i^{ask} - \overline{p}_i^{bid}}{0.5(p_i^{ask} + \overline{p}_i^{bid})}. \tag{C.14}$$

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.

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

C.5 Expected Prices and Liquidity from 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 (C.8), if dealer d's client is a k-type buyer. The weighted average price across all buyers of dealer d is

$$P_i^{ask} = E_d^b[\hat{P}_{i,d}(k)]. \tag{C.15}$$

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

$$P_i^{bid} = E_i^b[\hat{P}_{d,i}(k)]. \tag{C.16}$$

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

$$\Phi_i = \frac{P_i^{ask} - P_i^{bid}}{0.5P_i^{ask} + 0.5P_i^{bid}}. \tag{C.17}$$

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

C.6 Choosing Over Dealers

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,

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

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

$$\begin{aligned} V_i^o(k) &= \frac{\delta + kV_i^s}{r+k}. \\ &= \frac{r}{r+k}\left(\frac{\delta}{r}\right) + \frac{k}{r+k}\left(V_i^s\right) \end{aligned} \tag{C.18}$$

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:

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

From here,

$$\begin{aligned} V_i^s &= \frac{\delta - x + m_i^b\overline{p}_i^{bid}}{(r+m_i^b)} \\ &= \frac{r}{(r+m_i^b)}\left(\frac{\delta - x}{r}\right) + \frac{m_i^b}{(r+m_i^b)}\left(\overline{p}_i^{bid}\right). \end{aligned} \tag{C.19}$$

Thus, the seller's value function is a weighted average between the value of holding the bond forever, $\frac{\delta-x}{r}$, and the expected revenue from selling it, $\overline{p}_i^{bid}$.[35] 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.[36]

Thus, from (C.18), 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 (C.19), 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

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

[36]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.

value functions into the buyer's:

$$
\begin{aligned}
&V_i^b(k) = \\
&= \frac{m_i^s}{r+k+m_i^s}\left(V_i^o(k) - \overline{p}_i^{ask}(k)\right) \\
&= \frac{m_i^s}{r+k+m_i^s}\left[\frac{r}{r+k}\frac{\delta}{r} + \frac{k}{r+k}V_i^s - \overline{p}_i^{ask}(k)\right] \\
&= \frac{m_i^s}{r+k+m_i^s}\left(\frac{r}{r+k}\frac{\delta}{r} + \frac{k}{r+k}\left[\frac{r}{(r+m_i^b)}\frac{\delta-x}{r} + \frac{m_i^b}{(r+m_i^b)}\overline{p}_i^{bid}\right] - \overline{p}_i^{ask}(k)\right) \\
&= \frac{m_i^s}{r+k+m_i^s}\left(\frac{r}{r+k}\frac{\delta}{r} + \frac{k}{(r+k)}\frac{r}{(r+m_i^b)}\frac{\delta-x}{r} - \left[\overline{p}_i^{ask}(k) - \frac{k}{(r+k)}\frac{m_i^b}{(r+m_i^b)}\overline{p}_i^{bid}\right]\right) \\
&= \frac{m_i^s}{r+k+m_i^s}\left(\frac{r}{r+k}\frac{\delta}{r} + \frac{k}{(r+k)}\frac{r}{(r+m_i^b)}\frac{\delta-x}{r} - \phi^{eff}(k)\right),
\end{aligned}
\tag{C.20}
$$

where

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

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.[37] 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 $\lambda_{\text{DD}} > \lambda_{\text{D}}$, 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 $\lambda_{\text{DD}} = \lambda_{\text{D}}$, 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 $\lambda_{\text{DD}} < \lambda_{\text{D}}$, 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 $\lambda_{\text{DD}} > \lambda_{\text{D}}$), it already starts off narrower in absolute levels.

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

D Proof of Lemma 1

Proof of Lemma 1. We start by characterizing further $\hat{V}_i^b(k)$ and V_i^s for dealer $i \in N$. Consider first $\hat{V}_i^b(k)$. Using the ask-prices, the value function of a buyer who is a client of dealer $i \in N$ is given by

$$
\begin{aligned}
(r+k)\hat{V}_i^b(k) &= u\mu_i^s\left(\hat{V}_i^{\mathrm{ob}}(k) - V_i^s\right) + u_I\sum_{j\in N_i}\mu_j^s\left(\hat{V}_i^{\mathrm{ob}}(k) - V_j^s\right)\\
&= (u-u_I)\mu_i^s\left(\hat{V}_i^{\mathrm{ob}}(k) - V_i^s\right) + u_I\sum_{j\in N}\mu_j^s\left(\hat{V}_i^{\mathrm{ob}}(k) - V_j^s\right), \qquad \text{(D.1)}
\end{aligned}
$$

where N_i is the set all of dealers N except dealer i,

$$\hat{V}_i^{\mathrm{ob}}(k) \equiv \hat{V}_i^{\mathrm{o}}(k) - \hat{V}_i^{\mathrm{b}}(k)$$

is the reservation value of a buyer,

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

and

$$u \equiv \lambda_{\mathrm{D}} z_{\mathrm{D}}.$$

Combining the buyer and owner value functions, the reservation value of a buyer is

$$\hat{V}_i^{\mathrm{ob}}(k) = \frac{\delta + kV_i^s - (u_I-u)V_i^s\mu_i^s + u_I V_N^s}{k + r - (u_I-u)\mu_i^s + u_I\mu_N^s}, \qquad \text{(D.2)}$$

where

$$V_N^s \equiv \sum_{j\in N}\mu_j^s V_j^s$$

Using (D.2) into (D.1), the buyer's value function can be characterized as

$$\hat{V}_i^b(k) = \frac{1}{r+k}\frac{-(k+r)u_I V_N^s + u_I\left(kV_i^s+\delta\right)\mu_N^s - (u_I-u)\left(\delta - rV_i^s\right)\mu_i^s}{k + r - (u_I-u)\mu_i^s + u_I\mu_N^s} \qquad \text{(D.3)}$$

Consider next the value function of a seller client. Using the bid-price (C.7), the value function of a seller is

$$rV_i^s = \delta - x + u\left(\int_{\underline{k}}^{\overline{k}}\hat{\mu}_i^b(k)\left(\hat{V}_i^{\mathrm{ob}}(k) - V_i^s\right)dk\right) + u_I\sum_{j\in N_i}\left(\int_{\underline{k}}^{\overline{k}}\hat{\mu}_j^b(k)\left(\hat{V}_j^{\mathrm{ob}}(k) - V_i^s\right)dk\right). \qquad \text{(D.4)}$$

Integrating the reservation value of a buyer (D.2) over the buyer's mass

$$
\begin{aligned}
\int_{\underline{k}}^{\overline{k}}\hat{V}_i^{\mathrm{ob}}(k)\hat{\mu}_i^b(k)dk &= \int_{\underline{k}}^{\overline{k}}\frac{\delta + kV_i^s - (u_I-u)V_i^s\mu_i^s + u_I V_N^s}{k + r + (u-u_I)\mu_i^s + u_I\mu_N^s}\hat{\mu}_i^b(k)dk\\
&= V_i^s g_i^k + (\delta - (u_I-u)V_i^s\mu_i^s + u_I V_N^s)g_i\\
&= V_i^s(\mu_i^b - (r + (u-u_I)\mu_i^s + u_I\mu_N^s)g_i) + (\delta - (u_I-u)V_i^s\mu_i^s + u_I V_N^s)g_i\\
&= V_i^s(\mu_i^b - (r + u_I\mu_N^s)g_i) + (\delta + u_I V_N^s)g_i
\end{aligned}
$$

where

$$g_i \equiv \int_{\underline{k}}^{\overline{k}}\frac{1}{k + r + (u-u_I)\mu_i^s + u_I\mu_N^s}\hat{\mu}_i^b(k)dk$$

$$g_i^k \equiv \int_{\underline{k}}^{\overline{k}} \frac{k}{k+r+(u-u_I)\mu_i^s + u_I \mu_N^s} \hat{\mu}_i^b(k)dk$$

Thus,

$$\int_{\underline{k}}^{\overline{k}} V_i^{ob}(k)\hat{\mu}_i^b(k)dk = V_i^s(\mu_i^b - (r+u_I\mu_N^s)g_i) + (\delta + u_I V_N^s)g_i \tag{D.5}$$

Using (D.5), (D.4) becomes

$$\begin{aligned}
&rV_i^s = \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(D.6)}\\
&= \delta - x - (u_I - u)\int_{\underline{k}}^{\overline{k}} V_i^{ob}(k)\hat{\mu}_i^b(k)dk + (u_I - u)\mu_i^b V_i^s + u_I V_N^{ob} - u_I \mu_N^b V_i^s\\
&= \delta - x - (u_I - u)[V_i^s(\mu_i^b - (r + u_I\mu_N^s)g_i) + (\delta + u_I V_N^s)g_i] + (u_I - u)\mu_i^b V_i^s + u_I V_N^{ob} - u_I\mu_N^b V_i^s\\
&= \delta - x - (u_I - u)[-V_i^s(r + u_I\mu_N^s)g_i + (\delta + u_I V_N^s)g_i] + u_I V_N^{ob} - u_I\mu_N^b V_i^s
\end{aligned}$$

Thus, rearranging we get

$$V_i^s = \frac{\delta - x - (u_I - u)(\delta + u_I V_N^s)g_i + u_I V_N^{ob}}{r + u_I\mu_N^b - (u_I - u)(r + u_I\mu_N^s)g_i}, \tag{D.7}$$

where

$$V_N^{ob} \equiv \sum_{j\in N}\left(\int_{\underline{k}}^{\overline{k}} \hat{\mu}_j^b(k)V_j^{ob}(k)dk\right).$$

Next, consider the relationship between V_i^s and V_j^s for any two dealers i and j in N. Using (D.3),

$$(r+k)\left(\hat{V}_j^b(k) - \hat{V}_i^b(k)\right) = \frac{(k+r)(u_I - u)(u_I V_N^s + \delta)\left(\mu_i^s - \mu_j^s\right) - V_i^s p_j w_i + V_j^s p_i w_j}{p_i p_j}, \tag{D.8}$$

where

$$w_i \equiv r(u_I - u)\mu_i^s + k u_I \mu_N^s.$$

$$p_i \equiv k + r + (u - u_I)\mu_i^s + u_I\mu_N^s.$$

Then, at k^* such that $\hat{V}_i^b(k^*) = \hat{V}_j^b(k^*)$, the numerator of (D.8) is zero. Setting the numerator to zero and solving for V_j^s from it, we get:

$$V_j^s = \frac{(k^*+r)(u_I - u)(u_I V_N^s + \delta)\left(\mu_j^s - \mu_i^s\right) + V_i^s p_j^* w_i^*}{p_i^* w_j^*}, \tag{D.9}$$

where

$$w_i^* \equiv r(u_I - u)\mu_i^s + k^* u_I\mu_N^s.$$

$$p_i^* \equiv k^* + r + (u - u_I)\mu_i^s + u_I\mu_N^s.$$

Moreover, using (D.9),

$$\frac{V_j^s - V_i^s}{k^* + r} = \frac{\left(\mu_j^s - \mu_i^s\right)(u_I - u)\left((u_I V_N^s + \delta) - (r + u_I\mu_N^s)V_i^s\right)}{p_i^* w_j^*}. \tag{D.10}$$

Next, plug the expression for V_j^s from (D.9) into (D.8), simplify, and get

$$(r+k)\left(\hat{V}_j^b(k)-\hat{V}_i^b(k)\right)=$$
$$\left[(k^*+r)\,p_iw_j-(k+r)p_i^*w_j^*\right]\frac{\left(\mu_j^s-\mu_i^s\right)(u_I-u)\left((u_IV_N^s+\delta)-(r+u_I\mu_N^s)\,V_i^s\right)}{p_i^*w_j^*}\frac{1}{p_ip_j}$$

Using (D.10), we can express it as

$$(r+k)\left(\hat{V}_j^b(k)-\hat{V}_i^b(k)\right)=\frac{\left[(k^*+r)\,p_iw_j-(k+r)p_i^*w_j^*\right]}{(k^*+r)\,p_ip_j}\left(V_j^s-V_i^s\right) \tag{D.11}$$

Using the definitions of p_i, w_j, p_i^*, and w_j^* and rearranging,

$$\begin{aligned}&(r+k)\left(\hat{V}_j^b(k)-\hat{V}_i^b(k)\right)\\&=(k-k^*)\left(V_j^s-V_i^s\right)\\&\times\frac{\left[(k+r)\,(k^*+r)\,u_I\mu_N^s+ru_I^2\mu_N^s\left(\mu_N^s-\mu_i^s-\mu_j^s\right)+r\,(u-u_I)^2\mu_i^s\mu_j^s+ruu_I\mu_N^s\left(\mu_i^s+\mu_j^s\right)\right]}{(k^*+r)\,p_ip_j}\end{aligned} \tag{D.12}$$

The expression in square brackets is positive. Thus, the sign of $\hat{V}_j^b(k)-\hat{V}_i^b(k)$ is the same sign as that of $(k-k^*)\left(V_j^s-V_i^s\right)$.

This sorting mechanism of clients into dealers applies to any pair of dealers i and j even if their equilibrium client sets are not connected to each other on $[\underline{k},\overline{k}]$. That is, k^* does not have to correspond to one of the equilibrium cutoffs described in Theorem 1. □

E Existence Proofs

The asymmetric and symmetric equilibria are special cases of a more general equilibrium that we refer to as a semi-asymmetric equilibrium. The semi-asymmetric equilibrium features a tiered dealer structure. Let $V^s_{\tau,i}$ denote the liquidity quality of dealer i in tier τ and Ω_τ the set of dealers in tier τ. Then, we define a tier as a group of dealers such that $V^s_{\tau,i} = V^s_{\tau,j}$ for all i and j in Ω_τ. Analogous to the asymmetric equilibrium, let us index tiers in the order of increasing liquidity so that $V^s_{\tau,i} < V^s_{\tau+1,j}$ for all $i \in \Omega_\tau$ and $j \in \Omega_{\tau+1}$. Thus, dealers are homogenous within tiers but are heterogenous across tiers. Proposition E.1 shows that taking the number of tiers (denoted by η) and the number of dealers within a tier ($\{n_\tau\}_\tau$) as given, such equilibrium exists. Then, the symmetric equilibrium is a special case with a single tier ($\eta = 1$) and n dealers in that tier, while the asymmetric equilibrium is a special case with n tiers and a single dealer in each tier.

The within tier economy is analogous to the economy under the symmetric equilibrium. Dealers in the same tier offer identical liquidity ($V^s_{\tau,i}$) and prices ($\overline{p}^{bid}_i$, $\hat{p}^{ask}_i(k)$), have client masses of identical size, and serve the same set of investors. Investors are, in turn, indifferent between dealers in the same tier.

The dealer heterogeneity across tiers is analogous to the dealer heterogeneity in the asymmetric equilibrium. First, dealers in different tiers specialize in different investors. Those in tiers in which dealers offer high liquidity attract relatively high switching rate investors, while those in tiers with low liquidity dealers attract relatively buy-and-hold investors. When choosing dealers in different tiers, investors face the same tradeoff as in Proposition 1. Second, dealers in high liquidity tiers receive larger volumes of client orders than dealers in lower tiers. Third, dealers in different tiers have different network centrality. Dealers in high liquidity tiers trade more with other dealers, account for a larger fraction of the total interdealer volume, and form the core of the interdealer network. It is vice versa for dealers in low liquidity tiers. Thus, the results of the asymmetric equilibrium generalize to the semi-asymmetric equilibrium.

Proof of Lemma 2. The symmetric equilibrium is a special case of the semi-asymmetric equilibrium of Proposition E.1 with a single tier and n dealers in that tier. The existence of the symmetric equilibrium, as a result, is a corollary of Proposition E.1, which shows that semi-asymmetric equilibria exist for any number of tiers $\eta \leq n$ and any number of dealers in each tier $\{n_\tau\}_\tau$ and taking them as given.

A continuum of symmetric equilibria exist. To see this, the aggregate buyer mass is

$$\begin{aligned}
\mu^b_N &= \sum_{i\in N} \mu^b_i \\
&= \int_{\underline{k}}^{\overline{k}} \frac{f(k)}{k + (\lambda_{\text{D}} - \lambda_{\text{DD}})\frac{\mu^s_N}{n} + \lambda_{\text{DD}}\mu^s_N} \left[\sum_{i\in N} \nu_i(k)\right] dk \\
&= \int_{\underline{k}}^{\overline{k}} \frac{f(k)}{k + (\lambda_{\text{D}} - \lambda_{\text{DD}})\frac{\mu^s_N}{n} + \lambda_{\text{DD}}\mu^s_N} dk.
\end{aligned}$$

Thus, the aggregate buyer mass μ^b_N does not depend on the system of clients' dealer choices $\{\nu_{\tau,j}(k)\}$, but depends only on μ^s_N. Then, plugging this into the market clearing condition (C.3),

$$\int_{\underline{k}}^{\overline{k}} \frac{1}{k} f(k) dk - \int_{\underline{k}}^{\overline{k}} \frac{f(k)}{k + (\lambda_{\text{D}} - \lambda_{\text{DD}})\frac{\mu^s_N}{n} + \lambda_{\text{DD}}\mu^s_N} dk + \mu^s_N - S = 0. \tag{E.1}$$

The left-hand-side is negative at $\mu^s_N = 0$, strictly increasing in μ^s_N, and is equal to ∞ for $\mu^s_N = \infty$. Hence, it has a unique solution in μ^s_N. Although μ^s_N is uniquely determined, a continuum of dealer choice functions $\nu_i(k)$ satisfy (E.1). Thus, the multiplicity of the symmetric equilibrium is with respect to dealer choice functions, not client sizes. □

Proof of Theorem 1. The asymmetric equilibrium is a special case of the semi-asymmetric equilibrium of Proposition E.1 with $\eta = n$ tiers and a single dealer in each tier ($n_\tau = 1$). □

Proposition E.1 (Semi-Asymmetric Equilibrium)**.** *Suppose* $\{\Omega_\tau\}_\tau$*,* $\{n_\tau\}_{\tau=1}^{\tau=\eta}$*, and* η *are given, where* Ω_τ *is the set of dealers in tier* τ*,* n_τ *is the number of dealers in tier* τ*, and* η *is the number of tiers. An equilibrium exists in which* $V^s_{\tau,i} < V^s_{\tau+1,j}$ *for all* $i \in \Omega_\tau$*,* $j \in \Omega_{\tau+1}$*,* $\tau \in \{1, 2, \ldots, \eta - 1\}$*. It is characterized by cutoffs* $\{k^*_{1,2}, k^*_{2,3}, \ldots, k^*_{\eta-1,\eta}\}$*, where* $\underline{k} < k^*_{1,2} < \ldots < k^*_{\eta-1,\eta} < \overline{k}$*, buyers of type* $k < k^*_{1,2}$ *choose one of the dealers in tier 1, buyers of type* $k \in [k^*_{1,2}, k^*_{2,3}]$ *choose one of the dealers in tier 2, and so on. Buyers at the cutoff* $k = k^*_{\tau,\tau+1}$ *are indifferent between dealers in tiers* τ *and* $\tau + 1$*:* $V^b_{\tau,i}(k^*_{\tau,\tau+1}) = V^b_{\tau+1,j}(k^*_{\tau,\tau+1})$*.*[38] *In such equilibrium,* $\mu^s_{\tau,i} < \mu^s_{\tau+1,j}$ *and* $\mu^b_{\tau,i} < \mu^b_{\tau+1,j}$*. Within any tier* τ*,* $\mu^s_{\tau,i} = \mu^s_{\tau,j}$*,* $\mu^b_{\tau,i} = \mu^b_{\tau,j}$*,* $V^s_{\tau,i} = V^s_{\tau,j}$*,* $\overline{p}^{bid}_{\tau,i} = \overline{p}^{bid}_{\tau,j}$*,* $\overline{p}^{ask}_{\tau,i}(k) = \overline{p}^{ask}_{\tau,j}$ *for any two dealers* i *and* $j \in \Omega_\tau$*.*

Proof. We breakdown the proof into Lemmas E.1-E.4. In Lemma E.1, we show that Lemma 1 implies that (a) dealers with the same V^s_i specialize in the same continuous subset of clients on $[\underline{k}, \overline{k}]$ and (b) the client sets of dealers with different V^s_i's overlap at most at a single buyer type $k \in [\underline{k}, \overline{k}]$. Then, Lemmas E.2-E.4 show the rest of the results. □

Lemma E.1. *The client sets of dealers with different* V^s_i*'s overlap at most at a single buyer type* $k \in [\underline{k}, \overline{k}]$*. Dealers with the same* V^s_i *specialize in the same continuous subset of clients on* $[\underline{k}, \overline{k}]$*.*

Proof. These results are direct corollaries of Lemma 1.

First, suppose $V^s_i \neq V^s_j$ for dealers i and j, and k^* is such that $\hat{V}^b_j(k^*) = \hat{V}^b_i(k^*)$. From Lemma 1, the sign of $\hat{V}^b_j(k) - \hat{V}^b_i(k)$ is given by $(k - k^*)$ for all k. In turn, $(k - k^*) \neq 0$ for all $k \neq k^*$. This means that for all $k \neq k^*$, $\hat{V}^b_j(k) - \hat{V}^b_i(k) \neq 0$. Thus, the client sets of dealers with different V^s_i's overlap at a single buyer type $k \in [\underline{k}, \overline{k}]$.

Next, if $V^s_i = V^s_j$, then from Lemma 1, $\hat{V}^b_i(k) = \hat{V}^b_j(k)$ for all $k \in [\underline{k}, \overline{k}]$. Thus, each buyer $k \in [\underline{k}, \overline{k}]$ is indifferent between all dealers with the same V^s_i. □

Lemma E.2. *Let* $\tau \in \{1, 2, \ldots \eta\}$ *index dealer tiers and* Ω_τ *the set of dealers in tier* τ*. There exist* $\{k^*_{\tau,\tau+1}\}_{\tau=1}^{\tau=\eta-1}$ *such that for all tiers* $\tau \in \{1, 2, \ldots, \eta - 1\}$

(i) $\underline{k} < k^*_{1,2} < k^*_{2,3} < \ldots < k^*_{\eta-1,\eta} < \overline{k}$*,*

(ii) $\hat{V}^b_{\tau,i}(k^*_{\tau,\tau+1}) = \hat{V}^b_{\tau+1,j}(k^*_{\tau,\tau+1})$ *for all dealers* $i \in \Omega_\tau$ *and* $j \in \Omega_{\tau+1}$*,*

(iii) $\hat{V}^b_{\tau,i}(k) > \hat{V}^b_{\tau',i'}(k)$ *for all* $k \in (k^*_{\tau-1,\tau}, k^*_{\tau,\tau+1})$*,* $i \in \Omega_\tau$*,* $\tau' \neq \tau$*, and* $i' \in \Omega_{\tau'}$*, and*

(iv) $\hat{V}^b_{\tau,i}(k) = \hat{V}^b_{\tau,j}(k)$ *for all* $k \in (k^*_{\tau-1,\tau}, k^*_{\tau,\tau+1})$ *and* $i, j \in \Omega_\tau$*.*

Proof. Lemmas 1 and E.1 imply the following client structure across dealers. Consider the last two tiers: $\tau = \eta - 1$ and $\tau = \eta$. By definition, $V^s_{\eta-1,i} < V^s_{\eta,j}$ for all $i \in \Omega_{\eta-1}$ and $j \in \Omega_\eta$. If there is $k^*_{\eta-1,\eta}$ such that $\hat{V}^b_{\eta-1,i}(k^*_{\eta-1,\eta}) = \hat{V}^b_{\eta,j}(k^*_{\eta-1,\eta})$ for all for all $i \in \Omega_{\eta-1}$ and $j \in \Omega_\eta$, then using Lemma 1 buyers with $k > k^*_{\eta-1,\eta}$ choose one of the dealers in tier η. Moreover, because $V^s_{\eta,i} = V^s_{\eta,j}$ for all $i, j \in \Omega_\eta$, these buyers are indifferent between all the dealers in tier $\tau = \eta$. Buyers with switching rates below the cutoff $k < k^*_{\eta-1,\eta}$, on the other hand, prefer dealers in tier $\eta - 1$ over dealers in tier η.

It is analogous for any pair of neighboring dealers $\tau - 1$ and τ. For example, consider dealers in tiers $\tau = \eta - 2$ and $\tau = \eta - 1$. Again, by definition, $V^s_{\eta-2,i} < V^s_{\eta-1,j}$ for all $i \in \Omega_{\eta-2}$ and $j \in \Omega_{\eta-1}$. As a result, for some cutoff that we denote as $k^*_{\eta-2,\eta-1}$, all buyers with $k > k^*_{\eta-2,\eta-1}$ (but where $k < k^*_{\eta-1,\eta}$) choose one of the dealers in tier $\eta - 1$, while buyers with switching rates $k < k^*_{\eta-2,\eta-1}$ prefer dealers in tier $\eta - 2$ over dealers in tier $\eta - 1$.

[38] Proposition F.1 in the Appendix shows that the equilibrium is unique for a two-tier $\lambda_{DD} = \lambda_D$ environment.

Given such client structure, we can define cutoffs $\{k_{1,2}, k_{2,3}, ...k_{\eta-1,\eta}\}$ such that (if they exist) $\underline{k} \leq k^*_{1,2} \leq k^*_{2,3} \leq \ldots \leq k^*_{\eta-1,\eta} \leq \overline{k}$ and buyers of type $[k_{i-1,i}, k_{i,i+1}]$ choose one of the dealers in tier i. The remainder of this proof shows that such cutoffs exist.

Using (D.7) into (D.3), $(r+k)\left(\hat{V}^b_j(k) - \hat{V}^b_i(k)\right)$ for any two dealers in tiers i and j simplifies to

$$(r+k)\left(\hat{V}^b_j(k) - \hat{V}^b_i(k)\right) \tag{E.2}$$

$$= \frac{(u_I - u)\, e_{ij}(k) A}{p_i p_j \left(r + u_I \mu^b_N - (u_I - u)\left(r + u_I \mu^s_N\right) g_i\right)\left(r + u_I \mu^b_N - (u_I - u)\left(r + u_I \mu^s_N\right) g_j\right)}, \tag{E.3}$$

where

$$e_{i,j}(k)$$
$$\equiv (k+r)\left(\mu^s_j - \mu^s_i\right)\left(r + u_I \mu^b_N\right) + \left(r\,(u_I - u)\,\mu^s_i + k u_I \mu^s_N\right) p_j g_j - \left(r\,(u_I - u)\,\mu^s_j + k u_I \mu^s_N\right) p_i g_i$$

$$A \equiv u_I V^s_N \left(r + u_I \mu^b_N\right) + u_I \delta \left(\mu^b_N - \mu^s_N\right) + \left(x - u_I V^{\text{ob}}_N\right)\left(r + u_I \mu^s_N\right) \tag{E.4}$$

$$p_i \equiv k + r - (u_I - u)\mu^s_i + u_I \mu^s_N$$

$$p_j \equiv k + r - (u_I - u)\mu^s_j + u_I \mu^s_N$$

I show below in Lemma E.5 that $A \neq 0$. As a result, $k^*_{i,j}$ such that $\hat{V}^b_j(k^*_{i,j}) - \hat{V}^b_i(k^*_{i,j}) = 0$ is characterized by $e_{ij}(k^*_{i,j}) = 0$. Showing that the equilibrium exists amounts to showing that for all tiers $i \in \{1, 2, \ldots \eta - 1\}$ and its neighboring tier $j = i + 1$

$$\begin{aligned} e_{i,j}(k^*_{i,j}) =& (k^*_{i,j} + r)\left(\mu^s_i - \mu^s_j\right)\left(r + u_I \mu^b_N\right) \\ &+ \left(r\,(u_I - u)\,\mu^s_j + k^*_{i,j} u_I \mu^s_N\right) p_i g_i - \left(r\,(u_I - u)\,\mu^s_i + k^*_{i,j} u_I \mu^s_N\right) p_j g_j = 0 \end{aligned}$$

has a solution in $k^*_{i,j}$ where $V^s_j > V^s_i$. We illustrate the existence proof first for $\eta = 3$ tiers and then generalize the proof to a general number of tiers, η.

Suppose that there can be at most three tiers ($\eta = 3$). Consider the first cutoff $k^*_{1,2}$ characterized by $e_{1,2}(k^*_{1,2}) = 0$:

$$\begin{aligned} e_{1,2}(k^*_{1,2}) =& (k^*_{1,2} + r)\left(\mu^s_2 - \mu^s_1\right)\left(r + u_I \mu^b_N\right) \\ &+ \left(r\,(u_I - u)\,\mu^s_1 + k^*_{1,2} u_I \mu^s_N\right) p^*_2 g_2 - \left(r\,(u_I - u)\,\mu^s_2 + k^*_{1,2} u_I \mu^s_N\right) p^*_1 g_1 \\ =& 0 \end{aligned} \tag{E.5}$$

To show that there exists a cutoff $k^*_{1,2} \in (\underline{k}, \overline{k})$ satisfying $e_{1,2}(k^*_{1,2}) = 0$, it is sufficient to show that $e_{1,2}(k^*_{1,2} = \underline{k}) > 0$ and $e_{1,2}(k^*_{1,2} = \overline{k}) < 0$. If $k^*_{1,2} = \overline{k}$, by definition of the cutoffs, $k^*_{2,3} = \overline{k}$. Then, $\nu_1(k) > 0$ for all $k \in \left[\underline{k}, \overline{k}\right]$, $\mu^b_1 > 0$ and $g_1 > 0$, while $\nu_i(k) = 0$, $\mu^b_i = 0$, $g_i = 0$, for all $i \neq 1$. From the interdealer constraint (8), $\mu^s_i = \frac{\mu^s_N}{\mu^b_N}\mu^b_i$ for any dealer i. This implies that $\mu^s_1 > 0$ and $\mu^s_i = 0$ for $i \neq 1$. As a result, $e_{1,2}(k^*_{1,2} = \overline{k}) < 0$:

$$-(k^*_{1,2} + r)\mu^s_1 \left(r + u_I \mu^b_N\right) - k^*_{1,2} u_I \mu^s_N p^*_1 g_1$$

If instead $k^*_{1,2} = \underline{k}$, then the sign of $e_{12}(k^*_{1,2} = \underline{k})$ depends on whether $k_{2,3} > k_{1,2}$ when $k^*_{1,2} = \underline{k}$.

Whether $k_{2,3} \in (k_{1,2}, \overline{k}] = (\underline{k}, \overline{k}]$ depends on the sign of $e_{2,3}(k_{2,3})$ evaluated at $k_{2,3} = k_{1,2}$ and $k_{2,3} = \overline{k}$. If $k_{2,3} = \overline{k}$, since $k^*_{1,2} = \underline{k}$, we have that $g_i = 0$, $\mu^b_i = 0$, $\mu^s_i = 0$, for all dealers $i \neq 2$; thereby, $e_{2,3}(k_{2,3}) < 0$. If instead $k_{2,3} = k_{1,2} = \underline{k}$, $g_3 > 0$, $\mu^b_3 > 0$, $\mu^s_3 > 0$, while $g_2 = 0$, $\mu^b_2 = 0$,

$\mu_2^s = 0$. As a result, $e_{2,3}(k_{2,3} = \underline{k}) > 0$. Put together, when $k_{1,2}^* = \underline{k}$, there is $k_{2,3} \in (k_{1,2}, \overline{k})$ satisfying $e_{2,3}(k_{2,3}) = 0$.

The existence of $k_{2,3} \in (k_{1,2}, \overline{k}) = (\underline{k}, \overline{k})$ implies that $g_2 > 0$, $\mu_2^b > 0$, $\mu_2^s > 0$, $g_1 = 0$, $\mu_1^b = 0$, $\mu_1^s = 0$, and thereby $e_{1,2}(k_{1,2}^* = \underline{k}) > 0$ when $k_{1,2}^* = \underline{k}$. Thus, $e_{1,2}(k_{1,2}^* = \overline{k}) < 0$ and $e_{1,2}(k_{1,2}^* = \underline{k}) > 0$. By continuity, there exists a cutoff $k_{1,2}^* \in (\underline{k}, \overline{k})$ satisfying $e_{ij}(k_{1,2}^*) = 0$. Given the equilibrium cutoff $k_{1,2}^* \in (\underline{k}, \overline{k})$, there exists $k_{2,3}^* \in (k_{1,2}^*, \overline{k})$ (the argument is analogous to that of when $k_{1,2}^* = \underline{k}$). Thus, all three dealers attract positive mass of clients and exist in equilibrium.

We now show existence for general η number of tiers. Again, the first cutoff $k_{1,2}^*$ is characterized by $e_{1,2}(k_{1,2}^*) = 0$. To show that there exists a cutoff $k_{1,2}^* \in (\underline{k}, \overline{k})$ satisfying $e_{1,2}(k_{1,2}^*) = 0$, it is sufficient to show that $e_{1,2}(k_{1,2}^* = \underline{k}) > 0$ and $e_{1,2}(k_{1,2}^* = \overline{k}) < 0$.

If $k_{1,2}^* = \overline{k}$, by definition of the cutoffs, all the other cutoffs are also at the corner: $k_{i-1,i}^* = \overline{k}$. Then, $\nu_1(k) > 0$ for all $k \in \left[\underline{k}, \overline{k}\right]$, $g_1 > 0$, $\mu_1^s > 0$, while $\nu_i(k) = 0$, $g_i = 0$, $\mu_i^s = 0$ for all dealers i other than dealer 1: $i \neq 1$. As a result, $e_{ij}(k_{1,2}^* = \overline{k}) < 0$. If $k_{1,2}^* = \underline{k}$, then the sign of $e_{12}(k_{1,2}^* = \underline{k})$ depends on whether $k_{2,3} > k_{1,2}$ when $k_{1,2}^* = \underline{k}$.

Whether $k_{2,3} \in (k_{1,2}, \overline{k}]$ depends on the sign of $e_{2,3}(k_{2,3})$ evaluated at $k_{2,3} = k_{1,2}$ and $k_{2,3} = \overline{k}$. If $k_{2,3} = \overline{k}$, since $k_{1,2}^* = \underline{k}$, we have that $g_i = 0$, $\mu_i^b = 0$, $\mu_i^s = 0$, for all dealers $i \neq 2$; thereby, $e_{2,3}(k_{2,3}) < 0$. If instead $k_{2,3} = k_{1,2} = \underline{k}$, then analogous to the above, the sign of $e_{2,3}(k_{2,3})$ depends on whether $k_{3,4} > k_{2,3} = \underline{k}$. Iterating these arguments forward until the very last cutoff, $e_{i-1,i}(k_{i-1,i})$ evaluated at the left end of its feasible interval (i.e. at $k_{i,i-1} = k_{i-2,i-1}$) ultimately depends on the last cutoff $k_{\eta-1,\eta}$ when $k_{i-1,i} = \underline{k}$ for all $i < \eta$.

Consider then the last cutoff $k_{\eta-1,\eta}$ when all the previous cutoffs are $k_{i-1,i} = \underline{k}$. When $k_{\eta-1,\eta} = \underline{k}$, then $g_i = 0$, $\mu_i^b = 0$, $\mu_i^s = 0$ for all dealers $i < \eta$, while $g_\eta > 0$, $\mu_\eta^b > 0$, $\mu_\eta^s > 0$. As a result, $e_{\eta-1,\eta}(k_{\eta-1,\eta} = \underline{k}) > 0$. When instead $k_{\eta-1,\eta} = \overline{k}$, then $g_{\eta-1} > 0$, $\mu_{\eta-1}^b > 0$, $\mu_{\eta-1}^s > 0$, while $g_\eta = 0$, $\mu_\eta^b = 0$, $\mu_\eta^s = 0$. As a result, $e_{\eta-1,\eta}(k_{\eta-1,\eta} = \underline{k}) < 0$. Put together, when $k_{i-1,i} = \underline{k}$ for all $i < \eta$, there exists $k_{\eta-1,\eta}^* \in (\underline{k}, \overline{k})$ satisfying $e_{\eta-1,\eta}(k_{\eta-1,\eta}^*) = 0$.

The existence of $k_{\eta-1,\eta}^* \in (\underline{k}, \overline{k})$ in turn establishes the existence of $k_{\eta-2,\eta-1}^* \in (\underline{k}, \overline{k})$ when the cutoffs to its left are at the boundary, $\underline{k}$. To see this, the existence of $k_{\eta-1,\eta}^* \in (\underline{k}, \overline{k})$ implies that $g_{\eta-1} > 0$, $\mu_{\eta-1}^b > 0$, $\mu_{\eta-1}^s > 0$, $g_{\eta-2} = 0$, $\mu_{\eta-2}^b = 0$, $\mu_{\eta-2}^s = 0$ and thereby $e_{\eta-2,\eta-1}(k_{\eta-2,\eta-1} = \underline{k}) > 0$ when $k_{i-1,i} = \underline{k}$ for all $i < \eta$. If $k_{\eta-2,\eta-1} = \overline{k}$ while the previous cutoffs are at $\underline{k}$, then $g_{\eta-2} > 0$, $\mu_{\eta-2}^b > 0$, $\mu_{\eta-2}^s > 0$ and the masses of the other dealers are zero. As a result, $e_{\eta-2,\eta-1}(k_{\eta-2,\eta-1} = \overline{k}) < 0$. Put together, when $k_{1,2} = k_{2,3} = ... = k_{\eta-3,\eta-2} = \underline{k}$, there exists $k_{\eta-2,\eta-1} \in (\underline{k}, \overline{k})$.

We can iterate these results back to the first cutoff and show that treating the other cutoffs $\{k_{i-1,i}\}_{i>2}$ as implicit functions of $k_{1,2}$, there exists $k_{1,2}^* \in (\underline{k}, \overline{k})$ satisfying $e_{1,2}(k_{1,2}^* = \underline{k}) = 0$. Given $k_{1,2}^*$, there exists $k_{2,3}^* \in (k_{1,2}^*, \overline{k})$ satisfying $e_{2,3}(k_{2,3}^*) = 0$ because using the arguments above $e_{2,3}(k_{2,3} = \overline{k}) < 0$ and $e_{2,3}(k_{2,3} = \underline{k}) > 0$. Analogously, there exists $k_{3,4}^* \in (k_{1,2}^*, \overline{k})$ satisfying $e_{3,4}(k_{3,4}^*) = 0$ and so on. Thus, there exist cutoffs such that $\underline{k} < k_{1,2}^* < k_{2,3}^* < \ldots < k_{\eta-1,\eta}^* < \overline{k}$ and buyers of type $[k_{i-1,i}, k_{i,i+1}]$ choose one of the dealers in tier i. This implies that if we allow customers to sort across at most η tiers, all tiers exist in equilibrium. □

Lemma E.3. *$V_{\tau,i}^s = V_{\tau,j}^s$, $\mu_{\tau,j}^s = \mu_{\tau,i}^s$, $\mu_{\tau,j}^b = \mu_{\tau,j}^b$, $\overline{p}_{\tau,i}^{bid} = \overline{p}_{\tau,j}^{bid}$, $\overline{p}_{\tau,i}^{ask}(k) = \overline{p}_{\tau,j}^{ask}$, $M_{\tau,i}^{DD} = M_{\tau,j}^{DD}$ for any two dealers $i, j \in \Omega_\tau$ and for any tier τ.*

Proof. Consider the properties of dealers within a tier. By definition, dealers within a tier have identical V_τ^s. From Lemma E.1, any buyer $k \in [\underline{k}, \overline{k}]$ is indifferent between all dealers with the same V_i^s. From (D.10), $V_{\tau,i}^s = V_{\tau,j}^s$ for all $i, j \in \Omega_\tau$ implies that $\mu_{\tau,i}^s = \mu_{\tau,j}^s$ for all $i, j \in \Omega_\tau$. Since $\mu_i^b = \mu_i^s \frac{\mu_N^b}{\mu_N^s}$, this also implies that $\mu_{\tau,i}^b = \mu_{\tau,j}^b$ for all $i, j \in \Omega_\tau$. Thus, dealers in the same tier have identical client masses and serve the same segment of clients on $[\underline{k}, \overline{k}]$.

Although the masses are the same, the composition need not be. To see this, consider the buyer mass for a particular dealer in tier τ

$$\mu_\tau^b = \int_{k^*_{\tau-1,\tau}}^{k^*_{\tau,\tau+1}} \frac{\hat{f}(k)\nu_{\tau,i}(k)}{k+m_\tau^s} dk. \tag{E.6}$$

Suppose the equilibrium value is $\mu_\tau^b = a$ for some constant a. Fixing cutoffs $k^*_{\tau-1,\tau}$, $k^*_{\tau,\tau+1}$, and m_τ^s, a continuum of functions $\nu_{\tau,i}(k)$ exist such that the right-hand side equals a.

The additional result that $\overline{p}^{bid}_{\tau,i} = \overline{p}^{bid}_{\tau,j}$, $\overline{p}^{ask}_{\tau,i}(k) = \overline{p}^{ask}_{\tau,j}$, $M^{DD}_{\tau,i} = M^{DD}_{\tau,j}$ for any two dealers i and $j \in \Omega_\tau$ are direct implications from the fact buyer and seller value functions and client masses are identical across dealers within the same tier. □

Lemma E.4. *Suppose dealers in tier τ specialize in buyers with switching rates in $[k^*_{\tau-1,\tau}, k^*_{\tau,\tau+1}]$, while dealers in tier $\tau+1$ specialize in $[k^*_{\tau,\tau+1}, k^*_{\tau+1,\tau+2}]$. Then, $V^s_{\tau+1} > V^s_\tau$, $\mu^s_{\tau+1} > \mu^s_\tau$, and $\mu^b_{\tau+1} > \mu^b_\tau$, where the subscript denotes the tiers of the dealers.*

Proof. Lemma E.6 shows that $\delta - rV^s_\tau + u_I(V^s_N + V^s_\tau \mu^s_N) > 0$ for any dealer in tier τ. Using (D.10) and Assumption 10, $V^s_{\tau+1} > V^s_\tau$ if and only if $\mu^s_{\tau+1} > \mu^s_\tau$. Let us consider then $\mu^s_{\tau+1}$ and μ^s_τ.

At $k^*_{\tau,\tau+1}$ such that $\hat{V}^b_\tau(k^*_{\tau,\tau+1}) = \hat{V}^b_{\tau+1}(k^*_{\tau,\tau+1})$, it has to be that $e_{\tau,\tau+1}(k^*_{\tau,\tau+1}) = 0$. That is,

$$\begin{aligned} 0 &= (k^*_{\tau,\tau+1} + r)\left(\mu^s_\tau - \mu^s_{\tau+1}\right)\left(r + u_I \mu^b_N\right) \\ &\quad - \left(r\left(u_I - u\right)\mu^s_\tau + k^*_{\tau,\tau+1} u_I \mu^s_N\right) p_{\tau+1} g_{\tau+1} + \left(r\left(u_I - u\right)\mu^s_{\tau+1} + k^*_{\tau,\tau+1} u_I \mu^s_N\right) p_\tau g_\tau \\ &> (k^*_{\tau,\tau+1} + r)\left(\mu^s_\tau - \mu^s_{\tau+1}\right)\left(r + u_I \mu^b_N\right) \\ &\quad - \left(r\left(u_I - u\right)\mu^s_\tau + k^*_{\tau,\tau+1} u_I \mu^s_N\right) \mu^b_{\tau+1} + \left(r\left(u_I - u\right)\mu^s_{\tau+1} + k^*_{\tau,\tau+1} u_I \mu^s_N\right) \mu^b_\tau \\ &= (k^*_{\tau,\tau+1} + r)\left(\mu^s_\tau - \mu^s_{\tau+1}\right)\left(r + u_I \mu^b_N\right) - k^*_{\tau,\tau+1} u_I \mu^s_N \mu^b_{\tau+1} + k^*_{\tau,\tau+1} u_I \mu^s_N \mu^b_\tau \end{aligned} \tag{E.7}$$

The inequality uses the fact that

$$p_\tau g_\tau = \int_{k^*_{\tau-1,\tau}}^{k^*_{\tau,\tau+1}} \frac{k^*_{\tau,\tau+1} + r + (u - u_I)\mu^s_\tau + u_I \mu^s_N}{(k + r + (u - u_I)\mu^s_\tau + u_I \mu^s_N)} \hat{\mu}^b_\tau(k) dk > \int_{k^*_{\tau-1,\tau}}^{k^*_{\tau,\tau+1}} \hat{\mu}^b_\tau(k) dk = \mu^b_\tau$$

$$p_{\tau+1} g_{\tau+1} = \int_{k^*_{\tau,\tau+1}}^{k^*_{\tau+1,\tau+2}} \frac{k^*_{\tau,\tau+1} + r + (u - u_I)\mu^s_{\tau+1} + u_I \mu^s_N}{(k + r + (u - u_I)\mu^s_{\tau+1} + u_I \mu^s_N)} \hat{\mu}^b_{\tau+1}(k) dk < \int_{k^*_{\tau,\tau+1}}^{k^*_{\tau+1,\tau+2}} \hat{\mu}^b_{\tau+1}(k) dk = \mu^b_{\tau+1}$$

The equality in (E.7) uses $\mu^s_\tau \mu^b_{\tau+1} = \mu^s_{\tau+1} \mu^b_\tau$ (which is implied by the interdealer constraint (8)). The interdealer constraint also implies $\mu^b_{\tau+1} = \frac{\mu^b_N}{\mu^s_N} \mu^s_{\tau+1}$ and $\mu^b_\tau = \frac{\mu^b_N}{\mu^s_N} \mu^s_\tau$. Substituting these into the right-hand-side of (E.7), we get

$$0 > \left(\mu^s_\tau - \mu^s_{\tau+1}\right)\left[(k^*_{\tau,\tau+1} + r)\left(r + u_I \mu^b_N\right) + k^*_{\tau,\tau+1} u_I \mu^b_N\right]$$

The expression in square brackets is positive. Thus,

$$0 > \left(\mu^s_\tau - \mu^s_{\tau+1}\right).$$

In turn, $\mu^s_{\tau+1} > \mu^s_\tau$ implies $\mu^b_{\tau+1} > \mu^b_\tau$ since $\mu^b_i = \frac{\mu^b_N}{\mu^s_N} \mu^s_i$. □

Lemma E.5. $A \equiv u_I V^s_N \left(r + u_I \mu^b_N\right) + u_I \delta \left(\mu^b_N - \mu^s_N\right) + \left(x - u_I V^{ob}_N\right)\left(r + u_I \mu^s_N\right) \neq 0.$

Proof. We prove by contradiction. Suppose

$$A \equiv u_I V_N^s \left(r + u_I \mu_N^b\right) + u_I \delta \left(\mu_N^b - \mu_N^s\right) + \left(x - u_I V_N^{\text{ob}}\right)(r + u_I \mu_N^s) = 0. \tag{E.8}$$

From (D.7), $V_{\tau,i}^s$ differ across dealers only if $g_{\tau,i}$ differ across dealers. The derivative of the right-hand side of (D.7) with respect to $g_{\tau,i}$ is

$$\frac{(u - u_I)\left(u_I V_N^s \left(r + u_I \mu_N^b\right) + u_I \delta \left(\mu_N^b - \mu_N^s\right) + \left(x - u_I V_N^{\text{ob}}\right)(r + u_I \mu_N^s)\right)}{\left(r + g_{\tau,i} r u - g_{\tau,i} r u_I + u_I \mu_N^b + g_{\tau,i}(u - u_I) u_I \mu_N^s\right)^2}$$
$$= \frac{(u - u_I) A}{\left(r + g_{\tau,i} r u - g_{\tau,i} r u_I + u_I \mu_N^b + g_{\tau,i}(u - u_I) u_I \mu_N^s\right)^2}$$

Thus, since per our conjecture $A = 0$, $V_{\tau,i}^s$ do not differ across dealers (even if $g_{\tau,i}$ differ across dealers). Let $\hat{V}^s$ define the common $V_{\tau,i}^s$: $\hat{V}^s \equiv V_{\tau,i}^s$ for all τ and i. Using these implications, (D.7), and (D.5) and solving for $\hat{V}^s$, we get that for any dealer i in any tier τ

$$\hat{V}^s = \frac{\delta}{r} - \frac{1}{r} \frac{x}{1 + (u - u_I) g_{\tau,i} + u_I g_N}, \tag{E.9}$$

where $g_N \equiv \sum_{j \in N} g_j$. Eq. (E.9) shows that $g_{\tau,h}$ has to be the same across dealers also. Plugging (E.9) and (D.5) back into the right-hand-side of (E.8),

$$A = x \frac{r + \left[u_I \mu_N^b - (u_I - u)(r + u_I \mu_N^s) g_{\tau,i}\right]}{1 + (u - u_I) g_{\tau,i} + u_I g_N}$$

Consider the expression in square brackets. Since $g_{\tau,i}$ is the same across tiers, we can take the tier τ such that $\mu_\tau^s \leq \mu_{\tau'}^s$ for all τ' and express it as

$$\left[u_I \mu_N^b - (r + u_I \mu_N^s)(u_I - u) g_{\tau,i}\right]$$
$$= \mu_N^s \left[u_I \frac{\mu_N^b}{\mu_N^s} - \frac{1}{\mu_N^s}(r + u_I \mu_N^s)(u_I - u) g_{\tau,i}\right]$$
$$= \mu_N^s \left[u_I \frac{\mu_\tau^b}{\mu_\tau^s} - \frac{1}{\mu_N^s}(r + u_I \mu_N^s)(u_I - u) g_{\tau,i}\right]$$
$$= \frac{1}{\mu_\tau^s}\left[u_I \mu_\tau^b \mu_N^s - \mu_\tau^s (r + u_I \mu_N^s)(u_I - u) g_{\tau,i}\right]$$
$$= \frac{1}{\mu_\tau^s} \int_{\underline{k}}^{\overline{k}} \left(u_I \mu_N^s - \frac{\mu_\tau^s (r + u_I \mu_N^s)(u_I - u)}{k + r + (u - u_I)\mu_\tau^s + u_I \mu_N^s}\right) \frac{f(k) \nu_{\tau,i}(k)}{k + \lambda_{\text{D}} \mu_N^s} dk$$
$$= \frac{1}{\mu_\tau^s} \int_{\underline{k}}^{\overline{k}} \frac{k u_I \mu_N^s + r\left((u - u_I)\mu_\tau^s + u_I \mu_N^s\right) + u_I \mu_N^s \left(2u \mu_\tau^s + u_I (\mu_N^s - 2\mu_\tau^s)\right)}{k + r + (u - u_I)\mu_\tau^s + u_I \mu_N^s} \frac{f(k) \nu_{\tau,i}(k)}{k + \lambda_{\text{D}} \mu_N^s} dk$$

Since $\mu_N^s - 2\mu_\tau^s = (\mu_N^s - \mu_\tau^s) - \mu_\tau^s = \mu_{N_\tau}^s - \mu_\tau^s$ and $\mu_{\tau'}^s \geq \mu_\tau^s$ for all τ', we have that $u_I \mu_N^b - (r + u_I \mu_N^s)(u_I - u) g_{\tau,i} > 0$. Thus, A is strictly positive, and (E.8) is a contradiction. □

Lemma E.6. $\delta - r V_{\tau,i}^s + u_I (V_N^s + V_{\tau,i}^s \mu_N^s) \geq 0$ *for any dealer i in tier τ.*

Proof. Rearranging (D.7),

$$(u_I - u) g_{\tau,i} \left((\delta - r V_{\tau,i}^s) + u_I (V_N^s - \mu_N^s V_{\tau,i}^s)\right) = (\delta - x) - r V_{\tau,i}^s + u_I V_N^{\text{ob}} - V_{\tau,i}^s u_I \mu_N^b$$

Thus, the sign of $\delta - r V_{\tau,i}^s + u_I (V_N^s + V_{\tau,i}^s \mu_N^s)$ depends on the sign of $(\delta - x) - r V_{\tau,i}^s + u_I V_N^{\text{ob}} -$

$V^s_{\tau,i} u_I \mu^b_N$. Rearranging (D.6), we get

$$(u_I - u) \int_{\underline{k}}^{\overline{k}} \left(\hat{V}^{\text{ob}}_{\tau,i} - V^s_{\tau,i} \right) \hat{\mu}^b_{\tau,i}(k) = \delta - x - rV^s_{\tau,i} + u_I V^{\text{ob}}_N - u_I \mu^b_N V^s_{\tau,i}$$

On the left-hand-side, $\hat{V}^{\text{ob}}_{\tau,i} - V^s_{\tau,i}$ is the gains from trade, which is nonnegative. Thus, $\delta - rV^s_{\tau,i} + u_I(V^s_N + V^s_{\tau,i}\mu^s_N) \geq 0$. □

F Equilibrium Uniqueness

Proposition F.1. *Suppose $\lambda_{\text{D}} = \lambda_{\text{DD}}$, $\eta = 2$ is the number of tiers, $n_1 \geq 1$ and $n_2 \geq 1$ are the number of dealers in tiers 1 and 2. A unique semi-asymmetric equilibrium exists for any number of dealers in each tier.*

Proof. In a two-tier structure, the market clearing and interdealer conditions are

$$c = \left(n_1\mu_1^b + n_2\mu_2^b\right) - \left(n_1\mu_1^s + n_2\mu_2^s\right) \tag{F.1}$$

$$\mu_1^s\mu_2^b = \mu_2^s\mu_1^b \tag{F.2}$$

where $c \equiv \int_{\underline{k}}^{\bar{k}} f(k)\frac{1}{k}dk - S > 0$ is a constant and μ_i^b and μ_i^s denote the buyer and seller masses of an individual dealer in tier $i \in \{1, 2\}$. Solving for μ_1^b and μ_2^b from (F.1) and (F.2), we get:

$$\mu_1^b = \frac{\mu_1^s\left(c + n_1\mu_1^s + n_2\mu_2^s\right)}{n_1\mu_1^s + n_2\mu_2^s}$$

$$\mu_2^b = \frac{\mu_2^s\left(c + n_1\mu_1^s + n_2\mu_2^s\right)}{n_1\mu_1^s + n_2\mu_2^s}$$

Using the definitions of μ_1^b and μ_2^b, we have a system of two equations that characterize μ_1^s and μ_2^s as implicit functions of k^*.

$$\int_{\underline{k}}^{k^*} \frac{f(k)}{k + \lambda_{11}\mu_1^s + \lambda_{12}\mu_2^s}\nu_{1,i}(k)dk = \frac{\mu_1^s\left(c + n_1\mu_1^s + n_2\mu_2^s\right)}{n_1\mu_1^s + n_2\mu_2^s} \tag{F.3}$$

$$\int_{k^*}^{\overline{k}} \frac{f(k)}{k + \lambda_{21}\mu_1^s + \lambda_{22}\mu_2^s}\nu_{2,j}(k)dk = \frac{\mu_2^s\left(c + n_1\mu_1^s + n_2\mu_2^s\right)}{n_1\mu_1^s + n_2\mu_2^s} \tag{F.4}$$

where $c \equiv \int_{\underline{k}}^{\bar{k}} f(k)\frac{1}{k}dk - S > 0$, where $\lambda_{11} \equiv \lambda_{\text{D}} + (n_1 - 1)\,\lambda_{\text{DD}}$, $\lambda_{12} \equiv n_2\lambda_{\text{DD}}$, $\lambda_{21} \equiv n_1\lambda_{\text{DD}}$, $\lambda_{22} = \lambda_{\text{D}} + \lambda_{\text{DD}}\,(n_2 - 1)$. Then, μ_1^s, μ_2^s, and the cutoff k^* are the solution to a system of three equations: (F.3), (F.4), and

$$\begin{aligned} e \equiv &(k^* + r)\left(\mu_1^s - \mu_2^s\right)\left(r + u_I\mu_N^b\right) \\ &+ \left(r\left(u_I - u\right)\mu_1^s + k^*u_I\mu_N^s\right)p_2g_2 + \left(r\left(u_I - u\right)\mu_2^s + k^*u_I\mu_N^s\right)p_1g_1 = 0 \end{aligned} \tag{F.5}$$

where $p_i = r + k^* + (u - u_I) + u_I\mu_N^s$ for $i \in \{1, 2\}$. Equations (F.3) and (F.4) characterize μ_1^s and μ_2^s as implicit functions of k^*. Then, to show that (F.5) has a unique solution in k^*, it is sufficient to show that

$$\frac{\partial e(k^*, \mu_1^s, \mu_2^s)}{\partial k^*} = \frac{\partial e(k^*, \mu_1^s, \mu_2^s)}{\partial k^*} + \frac{\partial e(k^*, \mu_1^s, \mu_2^s)}{\partial \mu_1^s}\frac{\partial \mu_1^s(k^*)}{\partial k^*} + \frac{\partial e(k^*, \mu_1^s, \mu_2^s)}{\partial \mu_2^s}\frac{\partial \mu_2^s(k^*)}{\partial k^*} > 0 \tag{F.6}$$

In Lemmas F.1, F.2, and F.3, we show that $\frac{\partial e_0}{\partial k^*} > 0$, $\frac{\partial e_0}{\partial \mu_1^s} > 0$, and $\frac{\partial e_0}{\partial \mu_2^s} < 0$, respectively. Put together with $\frac{\partial \mu_1^s}{\partial k^*} > 0$ and $\frac{\partial \mu_2^s}{\partial k^*} < 0$ from Lemma F.4, we have (F.6). □

Lemma F.1. $\frac{\partial e}{\partial k^*} > 0$ *for any* λ_{D}, λ_{DD}, z, *and* z_{DD}.

Proof. Taking the partial derivative of the left-hand side of (F.5) with respect to k^*, we get

$$\frac{\partial e}{\partial k^*} = (\mu_1^s - \mu_2^s)\left(r + u_I\mu_N^b\right) + (u_I\mu_N^s)(p_1 g_1 - p_2 g_2) \tag{F.7}$$

$$+ (r(u_I - u)\mu_2^s + k^* u_I \mu_N^s) g_1 - (r(u_I - u)\mu_1^s + k^* u_I \mu_N^s) g_2 \tag{F.8}$$

$$+ (r(u_I - u)\mu_2^s + k^* u_I \mu_N^s) p_1 \frac{\partial g_1}{\partial k^*} - (r(u_I - u)\mu_1^s + k^* u_I \mu_N^s) p_2 \frac{\partial g_2}{\partial k^*} \tag{F.9}$$

where

$$\frac{\partial g_1}{\partial k^*} = \frac{f(k^*)\,\nu_{1,h}(k^*)}{r\left(k^* + (u - u_I)\mu_1^s + u_I\mu_N^s\right)\left(k^* + (\lambda_{\mathrm{D}} - \lambda_{\mathrm{DD}})\mu_1^s + \lambda_{\mathrm{DD}}\mu_N^s\right)}$$

$$\frac{\partial g_2}{\partial k^*} = -\frac{f(k^*)\,\nu_{2,h}(k^*)}{r\left(k^* + (u - u_I)\mu_2^s + u_I\mu_N^s\right)\left(k^* + (\lambda_{\mathrm{D}} - \lambda_{\mathrm{DD}})\mu_2^s + \lambda_{\mathrm{DD}}\mu_N^s\right)}$$

Since $\frac{\partial g_1}{\partial k^*} > 0$ and $\frac{\partial g_2}{\partial k^*} < 0$, the third row (F.9) is positive. Consider the combination of the first two rows (F.7) and (F.8) . From (F.5),

$$(\mu_1^s - \mu_2^s)\left(r + u_I\mu_N^b\right) = \frac{-(r(u_I - u)\mu_2^s + k^* u_I\mu_N^s)p_1 g_1 + (r(u_I - u)\mu_1^s + k^* u_I\mu_N^s)p_2 g_2}{r + k^*}$$

Substitute this into the first two rows and simplify to get:

$$\frac{(g_1 - g_2)\left(r(u - u_I)^2\mu_1^s\mu_2^s + u_I\left((k^*)^2 + 2k^* r + r(r + (u - u_I)(\mu_1^s + \mu_2^s))\right)\mu_N^s + r u_I^2(\mu_N^s)^2\right)}{k^* + r}$$

This is positive. Thus, $\frac{\partial e}{\partial k^*} > 0$. □

Lemma F.2. *Suppose* $\lambda_{\mathrm{D}} = \lambda_{\mathrm{DD}}$*, then* $\frac{\partial e}{\partial \mu_1^s} > 0$.

Proof. When $\lambda_{\mathrm{D}} = \lambda_{\mathrm{DD}}$, the aggregate seller mass μ_N^s does not depend on the system of dealer choices $\{\nu_{\tau,j}(k)\}$ for tiers $\tau \in \{1, 2\}$ and $j \in \Omega_\tau$. To see this, write the aggregate buyer mass as

$$\begin{aligned}
\mu_N^b &= \sum_{i\in\Omega_1}\mu_{1,i}^b + \sum_{i\in\Omega_2}\mu_{2,i}^b \\
&= \int_{\underline{k}}^{k^*} \frac{f(k)}{k + \lambda_{\mathrm{D}}\mu_N^s}\left[\sum_{i\in\Omega_1}\nu_{1,i}(k)\right]dk + \int_{k^*}^{\overline{k}} \frac{f(k)}{k + \lambda_{\mathrm{D}}\mu_N^s}\left[\sum_{i\in\Omega_2}\nu_{2,i}(k)\right]dk \\
&= \int_{\underline{k}}^{k^*} \frac{f(k)}{k + \lambda_{\mathrm{D}}\mu_N^s}dk + \int_{k^*}^{\overline{k}} \frac{f(k)}{k + \lambda_{\mathrm{D}}\mu_N^s}dk \\
&= \int_{\underline{k}}^{\overline{k}} \frac{f(k)}{k + \lambda_{\mathrm{D}}\mu_N^s}dk.
\end{aligned}$$

Thus, the aggregate buyer mass μ_N^b does not depend on the system of clients' dealer choices $\{\nu_{\tau,j}(k)\}$, but depends only on μ_N^s. Then, plugging this into the market clearing condition (C.3),

$$\int_{\underline{k}}^{\overline{k}} \frac{1}{k} f(k)dk - \int_{\underline{k}}^{\overline{k}} \frac{f(k)}{k + \lambda_{\mathrm{D}}\mu_N^s}dk + \mu_N^s = S, \tag{F.10}$$

(F.10) pins down μ_N^s. Since (F.10) does not depend on dealer choices $\{\nu_{\tau,j}(k)\}$, μ_N^s also does not depend on dealer choices. Derivatives of μ_N^b and μ_N^s with respect to the cutoff, k^*, as a result, are zero.

Taking the partial derivative of the left-hand side of (F.5) with respect to μ_1^s, we get

$$\frac{\partial e}{\partial \mu_1^s} = (r+k^*)\left(r+u_I\mu_N^b\right) - r\,(u_I-u)\,p_2 g_2 \tag{F.11}$$

$$+\left(r\,(u_I-u)\,\mu_2^s + k^* u_I \mu_N^s\right)(u_I-u)\left(-g_1 + p_1\frac{\partial g_1}{\partial \mu_1^s}\right) \tag{F.12}$$

where $\frac{\partial g_1}{\partial \mu_1^s} > 0$:

$$\begin{aligned}\frac{\partial g_1}{\partial \mu_1^s} &= \frac{\partial}{\partial \mu_1^s}\int_{\underline{k}}^{k^*} \frac{1}{k+r+(u-u_I)\,\mu_1^s + u_I\mu_N^s}\frac{f(k)\nu_{1,i}(k)}{k+\lambda_{\mathrm{D}}\mu_N^s}dk \\ &= (u_I-u)\int_{\underline{k}}^{k^*} \frac{1}{\left(k+r-(u_I-u)\,\mu_1^s+u_I\mu_N^s\right)^2}\frac{f(k)\nu_{1,i}(k)}{k+\lambda_{\mathrm{D}}\mu_N^s}dk\end{aligned}$$

Consider the first row, (F.11):

$$(r+k^*)\left(r+u_I\mu_N^b\right) - r\,(u_I-u)\,p_2g_2 = (r+k^*)\,r + k^*u_I\mu_N^b + ru_I\left(\mu_N^b - p_2g_2\right) + rup_2g_2.$$

To show that right-hand-side is positive, consider the expression $\mu_N^b - p_2g_2 > 0$:

$$\begin{aligned}\mu_N^b - p_2g_2 &> \mu_2^b - p_2g_2 \\ &= \int_{k^*}^{\overline{k}} \frac{f(k)\nu_{2,h}(k)}{k+\lambda_{\mathrm{D}}\mu_N^s}\,dk - p_2\int_{k^*}^{\overline{k}} \frac{f(k)\nu_{2,h}(k)}{\left(k+r+(u-u_I)\,\mu_2^s+u_I\mu_N^s\right)\left(k+\lambda_{\mathrm{D}}\mu_N^s\right)}\,dk \\ &= \int_{k^*}^{\overline{k}} \left(1-\frac{k^*+r+(u-u_I)\,\mu_2^s+u_I\mu_N^s}{k+r+(u-u_I)\,\mu_2^s+u_I\mu_N^s}\right)\frac{f(k)\nu_{2,h}(k)}{k+\lambda_{\mathrm{D}}\mu_N^s}\,dk.\end{aligned}$$

Since $k^* < k$, this is positive. Thus, (F.11) is positive.

To see the sign of the second row (F.12), consider $-g_1 + p_1\frac{\partial g_1}{\partial \mu_1^s}$:

$$\begin{aligned}&-g_1 + p_1\frac{\partial g_1}{\partial \mu_1^s} \\ &= -\int_{\underline{k}}^{k^*} \frac{f(k)\nu_{1,h}(k)}{\left(k+r-(u_I-u)\,\mu_1^s+u_i\mu_N^s\right)\left(k+\lambda_{\mathrm{D}}\mu_N^s\right)}dk \\ &\quad+\int_{\underline{k}}^{k^*} \frac{k^*+r-(u_I-u)\,\mu_1^s+u_I\mu_N^s}{k+r-(u_I-u)\,\mu_1^s+u_i\mu_N^s}\frac{f(k)\nu_{1,h}(k)}{\left(k+r-(u_I-u)\,\mu_1^s+u_I\mu_N^s\right)\left(k+\lambda_{\mathrm{D}}\mu_N^s\right)}dk \\ &= \int_{\underline{k}}^{k^*}\left(-1+\frac{k^*+r-(u_I-u)\,\mu_1^s+u_I\mu_N^s}{k+r-(u_I-u)\,\mu_1^s+u_I\mu_N^s}\right)\frac{f(k)\nu_{1,h}(k)}{\left(k+r-(u_I-u)\,\mu_1^s+u_I\mu_N^s\right)\left(k+\lambda_{\mathrm{D}}\mu_N^s\right)}dk \\ &= \int_{\underline{k}}^{k^*}\left(\frac{k^*-k}{k+r-(u_I-u)\,\mu_1^s+u_i\mu_N^s}\right)\frac{f(k)\nu_{1,h}(k)}{\left(k+r-(u_I-u)\,\mu_1^s+u_I\mu_N^s\right)\left(k+\lambda_{\mathrm{D}}\mu_N^s\right)}dk\end{aligned}$$

This is positive. Thus, $\frac{\partial e}{\partial \mu_1^s} > 0$. □

Lemma F.3. *Suppose* $\lambda_{\mathrm{D}} = \lambda_{\mathrm{DD}}$, *then* $\frac{\partial e}{\partial \mu_2^s} < 0$.

Proof. Taking the partial derivative of the left-hand side of (F.5) with respect to μ_2^s, we get

$$\frac{\partial e}{\partial \mu_2^s} = -\left(r+k^*\right)\left(r+u_I\mu_N^b\right) + r\left(u_I-u\right)p_1 g_1 + \left(r\left(u_I-u\right)\mu_1^s + k^* u_I \mu_N^s\right)\left(u_I-u\right)g_2 \quad \text{(F.13)}$$

$$-\left(r\left(u_I-u\right)\mu_1^s + k^* u_I\mu_N^s\right)\left(u_I-u\right)p_2\frac{\partial g_2}{\partial \mu_2^s}, \quad \text{(F.14)}$$

where $\frac{\partial g_2}{\partial \mu_2^s}$ is positive:

$$\begin{aligned}\frac{\partial g_2}{\partial \mu_2^s} &= \frac{\partial}{\partial \mu_2^s}\int_{k^*}^{\overline{k}} \frac{1}{k+r+\left(u-u_I\right)\mu_2^s + u_I\mu_N^s}\frac{f(k)\nu_{2,h}(k)}{k+\lambda_{\mathrm{D}}\mu_N^s}dk \\ &= \left(u_I-u\right)\int_{k^*}^{\overline{k}} \frac{1}{\left(k+r-\left(u_I-u\right)\mu_2^s + u_I\mu_N^s\right)^2}\frac{f(k)\nu_{2,h}(k)}{k+\lambda_{\mathrm{D}}\mu_N^s}dk.\end{aligned}$$

From (D.7), $V_2^s > V_1^s$ implies that $g_2 < g_1$. Then, using $g_2 < g_1$, the first row (F.13) is less than

$$\begin{aligned}&-\left(r+k^*\right)\left(r+u_I\mu_N^b\right) + r\left(u_I-u\right)p_1 g_1 + \left(r\left(u_I-u\right)\mu_1^s + k^* u_I\mu_N^s\right)\left(u_I-u\right)g_1 \\ &= -\left(r+k^*\right)\left[r+u_I\mu_N^b - \left(r+u_I\mu_N^s\right)\left(u_I-u\right)g_1\right]\end{aligned}$$

Consider now the sign of $\left[u_I\mu_N^b - \left(r+u_I\mu_N^s\right)\left(u_I-u\right)g_1\right]$

$$\begin{aligned}&\left[u_I\mu_N^b - \left(r+u_I\mu_N^s\right)\left(u_I-u\right)g_1\right] \\ &= \mu_N^s\left[u_I\frac{\mu_N^b}{\mu_N^s} - \frac{1}{\mu_N^s}\left(r+u_I\mu_N^s\right)\left(u_I-u\right)g_1\right] \\ &= \mu_N^s\left[u_I\frac{\mu_1^b}{\mu_1^s} - \frac{1}{\mu_N^s}\left(r+u_I\mu_N^s\right)\left(u_I-u\right)g_1\right] \\ &= \frac{1}{\mu_1^s}\left[u_I\mu_1^b\mu_N^s - \mu_1^s\left(r+u_I\mu_N^s\right)\left(u_I-u\right)g_1\right] \\ &= \frac{1}{\mu_1^s}\int_{\underline{k}}^{k^*}\left(u_I\mu_N^s - \frac{\mu_1^s\left(r+u_I\mu_N^s\right)\left(u_I-u\right)}{k+r+\left(u-u_I\right)\mu_1^s+u_I\mu_N^s}\right)\frac{f(k)\nu_{1,i}(k)}{k+\lambda_{\mathrm{D}}\mu_N^s}dk \\ &= \frac{1}{\mu_1^s}\int_{\underline{k}}^{k^*}\frac{ku_I\mu_N^s + r\left((u-u_I)\mu_1^s + u_I\mu_N^s\right) + u_I\mu_N^s\left(2u\mu_1^s + u_I\left((n_1-1)\mu_1^s + n_2\mu_2^s - \mu_1^s\right)\right)}{k+r+\left(u-u_I\right)\mu_1^s + u_I\mu_N^s}\frac{f(k)\nu_{1,i}(k)}{k+\lambda_{\mathrm{D}}\mu_N^s}dk\end{aligned}$$

Since $n_2 \geq 1$ and $\mu_2^s > \mu_1^s$, we have that $u_I\mu_N^b - \left(r+u_I\mu_N^s\right)\left(u_I-u\right)g_1 > 0$. Thus, the first row (F.13) is negative. The second row (F.14) is also negative. Put together, $\frac{\partial e_0}{\partial \mu_2^s} < 0$. □

Lemma F.4. *Suppose* $\lambda_{\mathrm{DD}} \geq \lambda_{\mathrm{D}}$. *Then,* $\frac{\partial \mu_1^s}{\partial k^*} > 0$ *and* $\frac{\partial \mu_2^s}{\partial k^*} < 0$.

Proof. Lemma C.1 showed that the following system of two equations characterize μ_1^s and μ_2^s as implicit functions of k^*.

$$\int_{\underline{k}}^{k^*}\frac{1}{k+\lambda_{11}\mu_1^s+\lambda_{12}\mu_2^s}\nu_1(k)dk = \frac{\mu_1^s\left(c+n_1\mu_1^s+n_2\mu_2^s\right)}{n_1\mu_1^s+n_2\mu_2^s} \quad \text{(F.15)}$$

$$\int_{k^*}^{\overline{k}}\frac{1}{k+\lambda_{21}\mu_1^s+\lambda_{22}\mu_2^s}\nu_2(k)dk = \frac{\mu_2^s\left(c+n_1\mu_1^s+n_2\mu_2^s\right)}{n_1\mu_1^s+n_2\mu_2^s} \quad \text{(F.16)}$$

where $\lambda_{11} \equiv \lambda_{\text{D}} + (n_1 - 1)\,\lambda_{\text{DD}}$, $\lambda_{12} \equiv n_2\lambda_{\text{DD}}$, $\lambda_{21} \equiv n_1\lambda_{\text{DD}}$, $\lambda_{22} = \lambda_{\text{D}} + \lambda_{\text{DD}}\,(n_2 - 1)$. I will use the following distribution $\nu_1(k) = \frac{1}{n_1}$ and $\nu_2(k) = \frac{1}{n_2}$.

Taking the derivative of both sides of (F.15) and (F.16) with respect to k^*, we get a linear system of equations

$$\left(\left(\mu_N^s\right)^2(1 + b_1\lambda_{11}) + cn_2\mu_2^s\right)\frac{\partial\mu_1^s}{\partial k^*} + \left(b_1\left(\mu_N^s\right)^2\lambda_{12} - cn_2\mu_1^s\right)\frac{\partial\mu_2^s}{\partial k^*} = f_1\left(\mu_N^s\right)^2 \tag{F.17}$$

$$\left(b_2\left(\mu_N^s\right)^2\lambda_{21} - cn_1\mu_2^s\right)\frac{\partial\mu_1^s}{\partial k^*} + \left(\left(\mu_N^s\right)^2(1 + b_2\lambda_{22}) + cn_1\mu_1^s\right)\frac{\partial\mu_2^s}{\partial k^*} = -f_2\left(\mu_N^s\right)^2 \tag{F.18}$$

where

$$f_1 \equiv \frac{f(k^*)\,\nu_{1,h}(k^*)}{k^* + m_1^s}$$

$$f_2 \equiv \frac{f(k^*)\,\nu_{2,h}(k^*)}{k^* + m_2^s}$$

$$b_1 \equiv \int_{\underline{k}}^{k^*}\frac{1}{\left(k + \lambda_{11}\mu_1^s + \lambda_{12}\mu_2^s\right)^2}\nu_1(k)dk \tag{F.19}$$

$$b_2 \equiv \int_{k^*}^{\overline{k}}\frac{1}{\left(k + \lambda_{21}\mu_1^s + \lambda_{22}\mu_2^s\right)^2}\nu_2(k)dk \tag{F.20}$$

Solving for $\frac{\partial\mu_1^s}{\partial k^*}$ and $\frac{\partial\mu_2^s}{\partial k^*}$ from (F.17) and (F.18), we get

$$\frac{\partial\mu_1^s}{\partial k^*} = \frac{\left((1 + b_2\lambda_{22})\left(\mu_N^s\right)^2 + cn_1\mu_1^s\right)\frac{\frac{f(k^*)}{n_1}}{k^*+m_1^s} + \left(b_1n_2\lambda_{\text{DD}}\left(\mu_N^s\right)^2 - cn_2\mu_1^s\right)\frac{\frac{f(k^*)}{n_2}}{k^*+m_2^s}}{D} \tag{F.21}$$

$$\frac{\partial\mu_2^s}{\partial k^*} = -\frac{\left((1 + b_1\lambda_{11})\left(\mu_N^s\right)^2 + cn_2\mu_2^s\right)\frac{\frac{f(k^*)}{n_2}}{k^*+m_2^s} + n_1\left(b_2\lambda_{\text{DD}}\left(\mu_N^s\right)^2 - c\mu_2^s\right)\frac{\frac{f(k^*)}{n_1}}{k^*+m_1^s}}{D} \tag{F.22}$$

where

$$\begin{aligned} D \equiv &\left(\left(\mu_N^s\right)^2(1 + b_1\lambda_{11}) + cn_2\mu_2^s\right)\left(\left(\mu_N^s\right)^2(1 + b_2\lambda_{22}) + cn_1\mu_1^s\right) \\ &- \left(b_1\left(\mu_N^s\right)^2\lambda_{12} - cn_2\mu_1^s\right)\left(b_2\left(\mu_N^s\right)^2\lambda_{21} - cn_1\mu_2^s\right). \end{aligned} \tag{F.23}$$

Consider next the numerator in (F.21). It can be expressed as $\frac{1}{f(k^*)}\left(k^* + m_1^s\right)\left(k^* + m_2^s\right)$ times

$$\lambda_{\text{DD}}\left(\mu_N^s\right)^2\left(k^* + m_1^s\right)b_1 + \frac{\left(\mu_N^s\right)^2\left(k^* + m_2^s\right)\lambda_{22}}{n_1}b_2 + \frac{\left(\mu_N^s\right)^2\left(k^* + m_2^s\right)}{n_1} - c\left(m_1^s - m_2^s\right)\mu_1^s \tag{F.24}$$

The last term in (F.24) can be written as $-c\left(m_1^s - m_2^s\right)\mu_1^s = -\mu_N^b m_1^s\mu_1^s + \mu_N^b m_2^s\mu_1^s + \mu_N^s\left(m_1^s - m_2^s\right)\mu_1^s$. Thus, all the terms except for $-\mu_N^b m_1^s\mu_1^s$ is positive. If we combine $-\mu_N^b m_1^s\mu_1^s$ with $\lambda_{\text{DD}}\left(\mu_N^s\right)^2\left(k^* + m_1^s\right)b_1$ in (F.24), we get

$$\begin{aligned} &\lambda_{\text{DD}}\left(\mu_N^s\right)^2\left(k^* + m_1^s\right)b_1 - m_1^s\mu_1^s\mu_N^b \\ &= \mu_N^s\left(\lambda_{\text{DD}}\mu_N^s\left(k^* + m_1^s\right)b_1 - m_1^s\mu_1^s\frac{\mu_N^b}{\mu_N^s}\right) \\ &= \mu_N^s\int_{\underline{k}}^{k^*}\frac{\lambda_{\text{DD}}\mu_N^s\left(k^* + m_1^s\right) - m_1^s\left(k + m_1^s\right)}{k + \lambda_{11}\mu_1^s + \lambda_{12}\mu_2^s}\frac{f(k)\nu_{1,h}(k)}{k + \lambda_{11}\mu_1^s + \lambda_{12}\mu_2^s}dk \end{aligned}$$

This is positive since $k^* \geq \underline{k}$ and, for $\lambda_{\text{DD}} \geq \lambda_{\text{D}}$, $\lambda_{\text{DD}}\mu_N^s \geq m_1^s$. Thus, combined with the result that $D > 0$ shown in Lemma F.5, $\frac{\partial \mu_1^s}{\partial k^*} > 0$.

Consider now the numerator in (F.22). It can be expressed as $\frac{1}{f(k^*)}\left(k^* + m_1^s\right)\left(k^* + m_2^s\right)$ times

$$-\left(\frac{\left(\mu_N^s\right)^2\left(k^* + m_1^s\right)\lambda_{11}}{n_2}b_1 + \left(\mu_N^s\right)^2\left(k^* + m_2^s\right)\lambda_{\text{DD}}b_2 + \frac{\left(\mu_N^s\right)^2\left(k^* + m_1^s\right)}{n_2} + c\left(m_1^s - m_2^s\right)\mu_2^s\right)$$

For $\lambda_{\text{DD}} \geq \lambda_{\text{D}}$, $m_1^s - m_2^s \geq 0$. Thus, together with the result $D > 0$ shown in Lemma F.5, $\frac{\partial \mu_1^s}{\partial k^*} < 0$. □

Lemma F.5. *$D > 0$, where D is given in* (F.23).

Proof. Dividing by $\left(\mu_N^s\right)^2$ and grouping terms with b_1 and b_2, $\frac{D}{\left(\mu_N^s\right)^2}$ is

$$\frac{D}{\left(\mu_N^s\right)^2} = \left(\mu_N^b n_1 m_1^s + \mu_N^s\left(\mu_N^s\lambda_{11} - n_1 m_1^s\right)\right)b_1 + \left(\mu_N^b n_2 m_2^s + \mu_N^s\left(\mu_N^s\lambda_{22} - n_2 m_2^s\right)\right)b_2$$
$$+ \left(\lambda_{22}\lambda_{11} - \lambda_{21}\lambda_{12}\right)\left(\mu_N^s\right)^2 b_1 b_2 + \mu_N^s\mu_N^b$$

where $m_1^s = \lambda_{11}\mu_1^s + \lambda_{12}\mu_2^s$ and $m_2^s = \lambda_{21}\mu_1^s + \lambda_{22}\mu_2^s$. Simplifying further and using the interdealer condition $\mu_i^b = \mu_i^s\frac{\mu_N^b}{\mu_N^s}$,

$$\frac{D}{\left(\mu_N^s\right)^2} = \tag{F.25}$$
$$\frac{\mu_N^s}{\mu_2^s}\left(\mu_2^b n_1 m_1^s + \frac{1}{2}\mu_2^s\left(\lambda_{22}\lambda_{11} - \lambda_{21}\lambda_{12}\right)\mu_N^s b_2\right)b_1$$
$$+ \frac{\mu_N^s}{\mu_1^s}\left(\mu_1^b n_2 m_2^s + \frac{1}{2}\mu_1^s\left(\lambda_{22}\lambda_{11} - \lambda_{21}\lambda_{12}\right)\mu_N^s b_1\right)b_2 \tag{F.26}$$
$$+ \mu_N^s\left(n_1\mu_1^b - \left(\lambda_{\text{DD}} - \lambda_{\text{D}}\right)n_2\mu_2^s b_1 + n_2\mu_2^b - \left(\lambda_{\text{DD}} - \lambda_{\text{D}}\right)n_1\mu_1^s b_2\right)$$

Using $\mu_1^b \equiv \int_{\underline{k}}^{k^*}\frac{1}{\left(k + \lambda_{11}\mu_1^s + \lambda_{12}\mu_2^s\right)}\nu_{1,h}(k)f(k)dk$ and $b_1 \equiv \int_{\underline{k}}^{k^*}\frac{1}{\left(k + \lambda_{11}\mu_1^s + \lambda_{12}\mu_2^s\right)^2}\nu_{1,h}(k)f(k)dk$, the expression $\mu_1^b n_2 m_2^s + \frac{1}{2}\mu_1^s\left(\lambda_{22}\lambda_{11} - \lambda_{21}\lambda_{12}\right)\mu_N^s b_1$ in (F.26) can be expressed as:

$$\mu_1^b n_2 m_2^s + \frac{1}{2}\mu_1^s\left(\lambda_{22}\lambda_{11} - \lambda_{21}\lambda_{12}\right)\mu_N^s b_1 =$$
$$= \int_{\underline{k}}^{k^*}\left(n_2 m_2^s - \frac{1}{2}\mu_1^s\mu_N^s\left(\lambda_{22}\lambda_{11} - \lambda_{21}\lambda_{12}\right)\frac{1}{k + m_1^s}\right)\frac{\nu_{1,h}(k)f(k)}{k + m_1^s}dk$$
$$= \int_{\underline{k}}^{k^*}\left(\frac{2m_1^s m_2^s n_2 + \left(-\lambda_{22}\lambda_{11} + \lambda_{21}\lambda_{12}\right)\mu_1^s\mu_N^s}{2\left(k + m_1^s\right)} + \frac{k m_2^s n_2}{k + m_1^s}\right)\frac{\nu_{1,h}(k)f(k)}{k + m_1^s}dk$$
$$= \frac{2m_1^s m_2^s n_2 + \left(-\lambda_{22}\lambda_{11} + \lambda_{21}\lambda_{12}\right)\mu_1^s\mu_N^s}{2}b_1 + \int_{\underline{k}}^{k^*}\frac{k m_2^s n_2}{k + m_1^s}\frac{\nu_{1,h}(k)f(k)}{k + m_1^s}dk$$

Substituting this back to (F.25), (F.25) becomes

$$\frac{D}{\left(\mu_N^s\right)^2} = \tag{F.27}$$
$$\left[\frac{\mu_2^b n_1 m_1^s + \frac{1}{2}\mu_2^s\left(\lambda_{22}\lambda_{11} - \lambda_{21}\lambda_{12}\right)\mu_N^s b_2}{\mu_2^s} + \frac{2m_1^s m_2^s n_2 + \left(\lambda_{21}\lambda_{12} - \lambda_{22}\lambda_{11}\right)\mu_1^s\mu_N^s}{2\mu_1^s} b_2\right] b_1$$
$$+ \frac{1}{\mu_1^s} b_2 \int_{\underline{k}}^{k^*} \frac{k m_2^s n_2}{k + m_1^s} \frac{f(k)}{k + \lambda_{\mathrm{D}}\mu_1^s + \lambda_{\mathrm{DD}}\mu_2^s} dk$$
$$+ n_1\mu_1^b - \left(\lambda_{\mathrm{DD}} - \lambda_{\mathrm{D}}\right) n_2\mu_2^s b_1 + n_2\mu_2^b - \left(\lambda_{\mathrm{DD}} - \lambda_{\mathrm{D}}\right) n_1\mu_1^s b_2$$

Consider the expression inside the square bracket in (F.27). Using the definitions of m_1^s, m_2^s, and μ_N^s, it simplifies to

$$\frac{n_1 m_1^s}{\mu_2^s}\mu_2^b + \frac{n_2\lambda_{21}\lambda_{11}\left(\mu_1^s\right)^2 + n_2\left(\lambda_{22}\lambda_{11} + \lambda_{21}\lambda_{12}\right)\mu_1^s\mu_2^s + n_2\lambda_{22}\lambda_{12}\left(\mu_2^s\right)^2}{\mu_1^s} b_2$$
$$= \int_{k^*}^{\overline{k}} \frac{\left(\lambda_{11}\mu_1^s + \lambda_{12}\mu_2^s\right)\left(k n_1\mu_1^s + \left(n_1\mu_1^s + n_2\mu_2^s\right)\left(\lambda_{21}\mu_1^s + \lambda_{22}\mu_2^s\right)\right)}{\mu_1^s\mu_2^s\left(k + \lambda_{21}\mu_1^s + \lambda_{22}\mu_2^s\right)^2}\nu_2(k)dk,$$

where the equality arises from using

$$\mu_2^b \equiv \int_{k^*}^{\overline{k}} \frac{1}{\left(k + \lambda_{21}\mu_1^s + \lambda_{22}\mu_2^s\right)}\nu_2(k)d$$

and

$$b_2 \equiv \int_{k^*}^{\overline{k}} \frac{1}{\left(k + \lambda_{21}\mu_1^s + \lambda_{22}\mu_2^s\right)^2}\nu_2(k)dk$$

and rearranging. Thus, the first row in (F.27) is strictly positive.

Consider next the third row of (F.27). Using $\mu_1^b \equiv \int_{\underline{k}}^{k^*} \frac{1}{\left(k+\lambda_{11}\mu_1^s+\lambda_{12}\mu_2^s\right)}\nu_{1,h}(k)dk$ and $b_1 \equiv \int_{\underline{k}}^{k^*} \frac{1}{\left(k+\lambda_{11}\mu_1^s+\lambda_{12}\mu_2^s\right)^2}\nu_{1,h}(k)dk$ and rearranging, $n_1\mu_1^b - \left(\lambda_{\mathrm{DD}} - \lambda_{\mathrm{D}}\right) n_2\mu_2^s b_1$ is

$$n_1\mu_1^b - \left(\lambda_{\mathrm{DD}} - \lambda_{\mathrm{D}}\right) n_2\mu_2^s b_1$$
$$= \int_{\underline{k}}^{k^*} \frac{k n_1 + n_1\left(\lambda_{\mathrm{D}} + \left(n_1 - 1\right)\lambda_{\mathrm{DD}}\right)\mu_1^s + n_2\left(\lambda_{\mathrm{D}} + \left(n_1 - 1\right)\lambda_{\mathrm{DD}}\right)\mu_2^s}{\left(k + \lambda_{11}\mu_1^s + \lambda_{12}\mu_2^s\right)^2}\nu_{1,h}(k)dk$$

Since $n_1 \geq 1$, $n_1\mu_1^b - \left(\lambda_{\mathrm{DD}} - \lambda_{\mathrm{D}}\right) n_2\mu_2^s b_1 > 0$. Analogously, using $\mu_2^b \equiv \int_{k^*}^{\overline{k}} \frac{1}{\left(k+\lambda_{21}\mu_1^s+\lambda_{22}\mu_2^s\right)}\nu_2(k)d$ and $b_2 \equiv \int_{k^*}^{\overline{k}} \frac{1}{\left(k+\lambda_{21}\mu_1^s+\lambda_{22}\mu_2^s\right)^2}\nu_2(k)dk$ and rearranging, $n_2\mu_2^b - \left(\lambda_{\mathrm{DD}} - \lambda_{\mathrm{D}}\right) n_1\mu_1^s b_2$ is

$$n_2\mu_2^b - \left(\lambda_{\mathrm{DD}} - \lambda_{\mathrm{D}}\right) n_1\mu_1^s b_2$$
$$= \int_{k^*}^{\overline{k}} \frac{k n_2 + n_1\left(\lambda_{\mathrm{D}} + \left(n_2 - 1\right)\lambda_{\mathrm{DD}}\right)\mu_1^s + n_2\left(\lambda_{\mathrm{D}} + \left(n_2 - 1\right)\lambda_{\mathrm{DD}}\right)\mu_2^s}{\left(k + \lambda_{21}\mu_1^s + \lambda_{22}\mu_2^s\right)^2}\nu_2(k)dk$$

This is also positive. Put together, D is strictly positive (and this is regardless of $\lambda_{\mathrm{DD}} \geq \lambda_{\mathrm{D}}$ or $\lambda_{\mathrm{DD}} < \lambda_{\mathrm{D}}$). □

G Proofs of Section 3-5 Results and Additional Discussion

Proof of Theorem 2. Consider two dealers indexed i and j where $\mu_i^s - \mu_j^s > 0$. Given that dealer i has a larger buyer and seller client mass than dealer j, it is straightforward to see that $M_i^D > M_j^D$.

Consider the interdealer volume, M_i^{DD}. Due to the interdealer constraints (8), 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 $\mu_i^b = \mu_i^s \frac{\mu_N^b}{\mu_N^s}$,

$$M_i^{DD} = 2\lambda_{\text{DD}}\mu_i^s \left(\sum_{j\in N_i} \mu_j^s\right) \frac{\mu_N^b}{\mu_N^s} = 2\lambda_{\text{DD}}\mu_i^s \mu_{N_i}^s \frac{\mu_N^b}{\mu_N^s}. \tag{G.1}$$

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

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

Since $\mu_i^s - \mu_j^s > 0$, we have that $M_i^{DD} - M_j^{DD} > 0$.

These results also apply to the semi-asymmetric equilibrium. Since $\mu_{\tau+1,i}^s > \mu_{\tau,j}^s$ for any dealer $i \in \Omega_{\tau+1}$ and $j \in \Omega_\tau$, it follows that $M_{\tau+1,i}^D > M_{\tau,j}^D$ and $M_{\tau+1,i}^{DD} > M_{\tau,j}^{DD}$. □

Proof of Proposition 1. This proposition highlights the cost associated with choosing a dealer that offers better liquidity. Consider the cost first for the parameter range $\lambda_{\text{D}} \geq \lambda_{\text{DD}}$. From (C.20),

$$V_\tau^b(k) = \frac{m_\tau^s}{r + k + m_\tau^s} \left[\frac{r}{r+k}\frac{\delta}{r} + \frac{k}{r+k} V_\tau^s - \bar{p}_\tau^{ask}(k)\right]$$

Solving for $(r+k)\bar{p}_\tau^{\text{ask}}(k)$, we get

$$\begin{aligned}
(r+k)\bar{p}_\tau^{\text{ask}}(k) &= -(r+k)\left(\frac{r+k}{m_\tau^s} + 1\right)\hat{V}_\tau^b(k) + kV_\tau^s + \delta \\
&\leq -(r+k)\left(\frac{r+k}{m_{\tau+1}^s} + 1\right)\hat{V}_\tau^b(k) + kV_\tau^s + \delta
\end{aligned}$$

The inequality uses the fact that $m_\tau^s \leq m_{\tau+1}^s$ when $\lambda_{\text{D}} \geq \lambda_{\text{DD}}$ (see below results on execution speed for more detail).

Taking the difference in prices, we get

$$\begin{aligned}
&(r+k)\left(\bar{p}_{\tau+1}^{\text{ask}}(k) - \bar{p}_\tau^{\text{ask}}(k)\right) \\
&\geq -(r+k)\left(\frac{r+k}{m_{\tau+1}^s} + 1\right)\hat{V}_{\tau+1}^b(k) + kV_{\tau+1}^s + (r+k)\left(\frac{r+k}{m_{\tau+1}^s} + 1\right)\hat{V}_\tau^b(k) - kV_\tau^s \\
&= -\left(\frac{r+k}{m_{\tau+1}^s} + 1\right)(r+k)\left(\hat{V}_{\tau+1}^b(k) - \hat{V}_\tau^b(k)\right) + k\left(V_{\tau+1}^s - V_\tau^s\right)
\end{aligned}$$

Using (D.11) and rearranging,

$$\begin{aligned}
&(r+k)\left(\bar{p}^{\text{ask}}_{\tau+1}(k)-\bar{p}^{\text{ask}}_{\tau}(k)\right)\\
&\geq \left[k-\left(\frac{r+k+m^s_{\tau+1}}{m^s_{\tau+1}}\right)\frac{\left[\left(k^*_{\tau,\tau+1}+r\right)p_\tau(k)w_{\tau+1}(k)-(k+r)p_\tau(k^*_{\tau,\tau+1})w_{\tau+1}(k^*_{\tau,\tau+1})\right]}{\left(k^*_{\tau,\tau+1}+r\right)p_\tau(k)p_{\tau+1}(k)}\right]\\
&\quad\times\left(V^s_{\tau+1}-V^s_\tau\right)\\
&=\frac{B}{m^s_{\tau+1}\left(k^*_{\tau,\tau+1}+r\right)p_\tau(k)p_{\tau+1}(k)}\left(V^s_{\tau+1}-V^s_\tau\right)
\end{aligned}$$

where

$$\begin{aligned}
B\equiv\ & km^s_{\tau+1}\left(k^*_{\tau,\tau+1}+r\right)p_\tau(k)p_{\tau+1}(k)\\
&-\left(r+k+m^s_{\tau+1}\right)\left[\left(k^*_{\tau,\tau+1}+r\right)p_\tau(k)w_{\tau+1}(k)-(k+r)p_\tau(k^*_{\tau,\tau+1})w_{\tau+1}(k^*_{\tau,\tau+1})\right]
\end{aligned}$$

At the limit $r\to 0$, B is $kk^*_{\tau,\tau+1}$ times

$$\begin{aligned}
&\left(u\mu^s_\tau+u_I\left(\mu^s_N-\mu^s_\tau\right)\right)\left(u\mu^s_{\tau+1}+u_I\left(\mu^s_N-\mu^s_{\tau+1}\right)\right)\left(\lambda_{\text{D}}\mu^s_{\tau+1}+\lambda_{\text{DD}}\left(\mu^s_N-\mu^s_{\tau+1}\right)\right)\\
&+k^*_{\tau,\tau+1}u_I\mu^s_N\left(\lambda_{\text{D}}\mu^s_{\tau+1}+\lambda_{\text{DD}}\left(\mu^s_N-\mu^s_{\tau+1}\right)\right)\\
&+k^2\left((\lambda_{\text{D}}-\lambda_{\text{DD}})\mu^s_{\tau+1}+(\lambda_{\text{DD}}-u_I)\mu^s_N\right)\\
&+k\left(k^*_{\tau,\tau+1}u_I\mu^s_N+\left(u\left(\mu^s_\tau+\mu^s_{\tau+1}\right)+u_I\left(\mu^s_N-\mu^s_\tau-\mu^s_{\tau+1}\right)\right)\left(\lambda_{\text{D}}\mu^s_{\tau+1}+\lambda_{\text{DD}}\left(\mu^s_N-\mu^s_{\tau+1}\right)\right)\right)
\end{aligned}$$

This is strictly positive. Since $V^s_{\tau+1}-V^s_\tau>0$, $\bar{p}^{\text{ask}}_{\tau+1}(k)-\bar{p}^{\text{ask}}_{\tau}(k)>0$ for all $k\in[\underline{k},\overline{k}]$ and $\tau\in\{1,2,...,\eta-1\}$.

If the parameter range is instead $\lambda_{\text{D}}<\lambda_{\text{DD}}$, then $\bar{p}^{\text{ask}}_{\tau+1}(k)-\bar{p}^{\text{ask}}_{\tau}(k)>0$ does not necessarily hold. The cost associated with choosing a high liquidity dealer is instead worse execution speed. To see this, from (19), the difference in the execution speed across any two dealers i and j is

$$m^s_i-m^s_j=-(\lambda_{\text{DD}}-\lambda_{\text{D}})\left(\mu^s_i-\mu^s_j\right)$$

from the perspective of a buyer client and

$$m^b_i-m^b_j=-(\lambda_{\text{DD}}-\lambda_{\text{D}})\left(\mu^b_i-\mu^b_j\right)$$

from the perspective of a seller client. Thus, when $\lambda_{\text{D}}<\lambda_{\text{DD}}$, the dealer with a larger client mass offers worse execution speed than a dealer with a smaller client mass. The dealers with larger client masses are the dealers with better overall liquidity, V^s_τ, (this is true regardless of the parameter values). Put together, the dealers with better overall liquidity offer worse execution speed. Thus, for the parameter range $\lambda_{\text{D}}<\lambda_{\text{DD}}$, buyers tradeoff execution speed versus overall liquidity. □

Proof of Proposition 2. Using (C.8)

$$\begin{aligned}
P^{ask}_{i,d}\equiv E^b_d\left[\hat{P}_{i,d}(k)\right]&=\frac{1}{2}V^s_i\int_{\underline{k}}^{\overline{k}}\frac{\hat{\mu}^b_j(k)}{\mu^b_j}+\frac{1}{2}\int_{\underline{k}}^{\overline{k}}\frac{\hat{\mu}^b_d(k)}{\mu^b_d}(V^o_d(k)-V^b_d(k)).\\
&=\frac{1}{2}V^s_i+\frac{1}{2}E^b_d\left[V^{ob}_d(k)\right].
\end{aligned}\tag{G.2}$$

Since $V^s_c>V^s_p$,

$$P^{ask}_{c,d}>P^{ask}_{p,d}.$$

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 arbitrary dealer d sells back to dealer i

$$P_{d,i}^{bid} = \frac{1}{2}V_d^s + \frac{1}{2}\int_{\underline{k}}^{\overline{k}} \frac{\hat{\mu}_i^b(k)}{\mu_b^b} V_i^{ob}(k). \tag{G.3}$$

$$= \frac{1}{2}V_d^s + \frac{1}{2}E_i^b V_i^{ob}. \tag{G.4}$$

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:

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

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

Proof of Proposition 3. Integrating the value functions over the respective client masses yields:

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

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

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

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

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

Simplifying it and replacing $\hat{\mu}_i^b(k)$ and $\hat{\mu}_i^o(k)$ with $\hat{\mu}_i^b(k) = \frac{\hat{f}(k)\nu_i(k)}{k+m_i^s}$ and $\hat{\mu}_i^o(k) = \frac{\hat{f}(k)\nu_i(k)m_i^s}{k\left(k+m_i^s\right)}$, we get

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

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

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

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:

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

where the second equality comes from the market clearing condition. □

G.1 Dealer Interconnectedness

Interconnectedness affects dealer profits and clients' welfare as follows. Because introducing the interdealer market lengthens the average intermediation chain, the effects are similar to the effects of market fragmentation discussed in Section 5.2. As Figure 8 illustrates, how clients and dealers split the increase in the welfare depends on clients' bargaining power in two-dealer chains (z_{DD}) relative to that in one-dealer chains (z_{D}). If it is significantly larger in two-dealer chains, clients not only extract the entire increase in the welfare but also get a cut from dealers' profit. That is, by lengthening the intermediation chain, clients tilt the gains from trade in their favor at the expense of dealers. Dealers in this case are better off without the interdealer market. For an intermediate range of clients' bargaining power in two-dealer chains, both clients and dealers benefit from interconnectedness. Finally, if clients' bargaining power is smaller in two-dealer chains than in one-dealer chains, now dealers extract the increase in the welfare. They do so at the expense of customer welfare.

Proof of Proposition 4. The environments with and without the interdealer market are the environments in which $\lambda_{\text{DD}} > 0$ and $\lambda_{\text{DD}} = 0$, respectively. Thus, to show the results of Proposition 4, showing how the total welfare and the volume of trade change with respect to λ_{DD} for all values of λ_{D} and λ_{DD} is sufficient. Moreover, we prove Proposition 4 results by comparing the symmetric equilibria with and without the interdealer market. The effect of interconnectedness across the asymmetric equilibria is qualitatively the same but more tedious to prove analytically.

Since the total welfare depends only on the aggregates mass of sellers, μ_N^s, consider μ_N^s. In the symmetric equilibrium, $\mu_i^s = \frac{\mu_N^s}{n}$ and $\mu_i^b = \frac{\mu_N^b}{n}$ for all $i \in N$. Applying (C.3) to the symmetric equilibrium, we get

$$
\int_{\underline{k}}^{\overline{k}} \frac{1}{k} f(k) dk - \int_{\underline{k}}^{\overline{k}} \frac{f(k)}{k + (\lambda_{\text{D}} + (n-1)\lambda_{\text{DD}})\,\frac{\mu_N^s}{n}} dk + \mu_N^s = S. \tag{G.5}
$$

Applying the Implicit Function Theorem to (G.5), the derivative of μ_N^s with respect to λ_{DD} is

$$\frac{\partial \mu_N^s}{\partial \lambda_{\text{DD}}} = -\frac{\frac{(n-1)\mu_N^s}{n}\int_{\underline{k}}^{\overline{k}} \frac{f(k)}{\left(k+(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}\right)^2}dk}{1+(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{1}{n}\int_{\underline{k}}^{\overline{k}} \frac{f(k)}{\left(k+(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}\right)^2}dk}. \tag{G.6}$$

The right-hand-side is negative for any λ_{DD} and λ_{D}. Thus, the aggregate mass of sellers decreases with the introduction of the interdealer market. Using (27), this implies that the total welfare increases with interconnectedness.

The aggregate volume is

$$\begin{aligned}\sum_{i\in N}\left(M_i^D + M_i^{DD}\right) &= n\mu_i^b\left(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}}\right)\mu_i^s \\ &= \left(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}}\right)\frac{\mu_N^s}{n}\int_{\underline{k}}^{\overline{k}}\frac{f(k)}{k+(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}}dk, \end{aligned}\tag{G.7}$$

where the second equality substitutes in $n\mu_i^b = \int_{k_1}^{k_2}\frac{f(k)}{k+(\lambda+(n-1)\lambda_i)\mu_1^s}dk$ and $\mu_i^s = \frac{\mu_N^s}{n}$. Then, taking the derivative with respect to λ_{DD} and simplifying, we get

$$\begin{aligned}&\frac{\partial \sum_{i\in N}\left(M_i^D + M_i^{DD}\right)}{\partial \lambda_{\text{DD}}} = \\ &= \frac{\partial\left(\left(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}}\right)\mu_N^s\right)}{\partial\lambda_{\text{DD}}}\left[\frac{1}{n}\int_{\underline{k}}^{\overline{k}}\frac{kf(k)}{\left(k+(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}\right)^2}dk\right].\end{aligned}\tag{G.8}$$

From (G.8), the sign depends on the derivative of the execution speed, $\frac{\partial\left((\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}\right)}{\partial\lambda_{\text{DD}}}$:

$$\frac{\partial\left(\left(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}}\right)\mu_N^s\right)}{\partial\lambda_{\text{DD}}} = \frac{\partial\mu_N^s}{\partial\lambda_{\text{DD}}}\left(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}}\right)+(n-1)\mu_N^s$$

Substituting in (G.6) and simplifying, we get

$$\frac{\partial\left(\left(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}}\right)\mu_N^s\right)}{\partial\lambda_{\text{DD}}} = \frac{(n-1)\mu_N^s}{1+(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{1}{n}\int_{\underline{k}}^{\overline{k}}\frac{f(k)}{\left(k+(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}\right)^2}dk}.$$

This is positive. Thus, both execution speeds and the aggregate volume of trade increase with interconnectedness. □

Proof of Proposition 5. 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,

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

Or, alternatively,

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

□

G.2 Market Fragmentation

Proof of Proposition 6. Applying the Implicit Function Theorem to (G.5) and simplifying, we get

$$\frac{\partial \mu_N^s}{\partial n} = -\frac{\frac{(\lambda_{\text{DD}}-\lambda_{\text{D}})}{n}\frac{\mu_N^s}{n}\int_{\underline{k}}^{\overline{k}} \frac{f(k)}{\left(k+\frac{(\lambda_{\text{DD}}+(n-1)\lambda_{\text{D}})\mu_N^s}{n}\right)^2}dk}{1+\int_{\underline{k}}^{\overline{k}} \frac{\frac{1}{n}(\lambda_{\text{DD}}+(n-1)\lambda_{\text{D}})f(k)}{\left(k+\frac{(\lambda_{\text{DD}}+(n-1)\lambda_{\text{D}})\mu_N^s}{n}\right)^2}dk} \tag{G.9}$$

Thus, the sign depends on $-(\lambda_{\text{DD}} - \lambda_{\text{D}})$. If, for example, $\lambda_{\text{DD}} - \lambda_{\text{D}} > 0$, it is negative implying that the total welfare increases in n.

Consider now how increasing the number of dealers affects the aggregate volume of trade. Taking the derivative of (G.7) with respect to n, we get

$$\frac{\partial}{\partial n}\sum_{i\in N}\left(M_i^D + M_i^{DD}\right) = \int_{\underline{k}}^{\overline{k}} \frac{kf(k)}{\left(k+(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}\right)^2}dk\frac{\partial}{\partial n}\left[(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}\right] \tag{G.10}$$

The sign depends on $\frac{\partial\left((\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}\right)}{\partial n}$:

$$\frac{\partial\left((\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}\right)}{\partial n} = \frac{\frac{\partial \mu_N^s}{\partial n}(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}}) + (\lambda_{\text{DD}}-\lambda_{\text{D}})\mu_N^s}{n^2}.$$

Substituting in (G.9) and simplifying, we get

$$\frac{\partial\left((\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\frac{\mu_N^s}{n}\right)}{\partial n} = \frac{1}{n^2}\frac{(\lambda_{\text{DD}}-\lambda_{\text{D}})\mu_N^s}{1+\int_{\underline{k}}^{\overline{k}} \frac{\frac{1}{n}(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})f(k)}{\left(k+\frac{(\lambda_{\text{D}}+(n-1)\lambda_{\text{DD}})\mu_N^s}{n}\right)^2}dk}.$$

Thus, the sign depends on $\lambda_{\text{DD}} - \lambda_{\text{D}}$ and is positive if $\lambda_{\text{DD}} - \lambda_{\text{D}} > 0$. □

Proof of Proposition 7. The welfare maximizing cutoff, k^*, is such that

$$\frac{\partial \mu_N^s}{\partial k^*} = \frac{\partial(\mu_1^s+\mu_2^s)}{\partial k^*} = \frac{\partial \mu_1^s}{\partial k^*} + \frac{\partial \mu_2^s}{\partial k^*} = 0.$$

Using (F.17) and (F.18) and simplifying, the welfare maximizing cutoff is characterized by:

$$\mu_N^s(k^*+\lambda_{\text{D}}\mu_1^s+\lambda_{\text{DD}}\mu_2^s)b_1 - \mu_N^s(k^*+\lambda_{\text{DD}}\mu_1^s+\lambda_{\text{D}}\mu_2^s)b_2 - (\mu_2^s-\mu_1^s)\mu_N^b = 0. \tag{G.11}$$

where b_1 and b_2 are given by (F.19) and (F.20). Divide (G.11) by μ_N^s and use the interdealer condition $\mu_i^b = \mu_i^s \frac{\mu_N^b}{\mu_N^s}$:

$$(k^* + \lambda_{\text{D}}\mu_1^s + \lambda_{\text{DD}}\mu_2^s)b_1 - (k^* + \lambda_{\text{DD}}\mu_1^s + \lambda_{\text{D}}\mu_2^s)b_2 - (\mu_2^b - \mu_1^b) = 0.$$

Rearranging it,

$$\begin{aligned}0 &= (k^* + \lambda_{\text{D}}\mu_1^s + \lambda_{\text{DD}}\mu_2^s)b_1 + \mu_1^b - (k^* + \lambda_{\text{DD}}\mu_1^s + \lambda_{\text{D}}\mu_2^s)b_2 - \mu_2^b \\ &> 2(\mu_1^b - \mu_2^b).\end{aligned}$$

To see how the inequality arises, consider $(k^* + \lambda_{\text{D}}\mu_1^s + \lambda_{\text{DD}}\mu_2^s)b_1 + \mu_1^b$:

$$(k^* + \lambda_{\text{D}}\mu_1^s + \lambda_{\text{DD}}\mu_2^s)b_1 \tag{G.12}$$

$$= \int_{\underline{k}}^{k^*} \frac{(k^* + \lambda_{\text{D}}\mu_1^s + \lambda_{\text{DD}}\mu_2^s)}{(k + \lambda_{\text{D}}\mu_1^s + \lambda_{\text{DD}}\mu_2^s)^2}\nu_1(k)dk \tag{G.13}$$

$$> \int_{\underline{k}}^{k^*} \frac{(k + \lambda_{\text{D}}\mu_1^s + \lambda_{\text{DD}}\mu_2^s)}{(k + \lambda_{\text{D}}\mu_1^s + \lambda_{\text{DD}}\mu_2^s)^2}\nu_1(k)dk \tag{G.14}$$

$$= \int_{\underline{k}}^{k^*} \frac{1}{(k + \lambda_{\text{D}}\mu_1^s + \lambda_{\text{DD}}\mu_2^s)}\nu_1(k)dk \tag{G.15}$$

$$= \mu_1^b. \tag{G.16}$$

Thus,

$$(k^* + \lambda_{\text{D}}\mu_1^s + \lambda_{\text{DD}}\mu_2^s)b_1 > \mu_1^b.$$

Analogously, $-(k^* + \lambda_{\text{DD}}\mu_1^s + \lambda_{\text{D}}\mu_2^s)b_2 > -\mu_2^b$. Then, $0 > \mu_1^b - \mu_2^b$ or $\mu_2^b > \mu_1^b$ at the welfare maximizing cutoff. □

References

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

Atkeson, Andrew, Andrea Eisfeldt, and Pierre-Olivier 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, Nicolae Gârleanu, and Lasse 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 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, Artem, 2012, Bid-ask spreads and the over-the-counter interdealer markets: Core and peripheral dealers, Working paper.

———, 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 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, 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.