Core-Periphery Trading Networks

Chaojun Wang*
Stanford University

Job Market Paper
February 10, 2017
Click here for the latest version

Abstract

Core-periphery trading networks arise endogenously in over-the-counter markets as an equilibrium balance between trade competition and inventory efficiency. A small number of firms emerge as core dealers to intermediate trades among a large number of peripheral firms. The equilibrium number of dealers depends on two countervailing forces: *(i)* competition among dealers in their pricing of immediacy to peripheral firms, and *(ii)* the benefits of concentrated intermediation for lowering dealer inventory risk through dealers' ability to quickly net purchases against sales. For an asset with a lower frequency of trade demand, intermediation is concentrated among fewer dealers, and interdealer trades account for a greater fraction of total trade volume. These two predictions are strongly supported by evidence from the Bund and U.S. corporate bond markets. From a welfare viewpoint, I show that there are too few dealers for assets with frequent trade demands, and too many for assets with infrequent trade demands.

JEL Classifications: C73, D43, D85, L13, L14, G14

Keywords: Over-the-counter, network, financial intermediation, trade competition, inventory cost

*Email: chaojunw@stanford.edu Web: chaojunw.web.stanford.edu I am deeply indebted to my advisors Darrell Duffie, Matthew Jackson and Arvind Krishnamurthy for their continued guidance and detailed suggestions. I am grateful for helpful discussions from Ana Babus, Svetlana Bryzgalova, Peter DeMarzo, Matthew Elliott, Joel Hasbrouck, Hanno Lustig, Albert Menkveld, Rémy Praz, Johan Walden, Yao Zeng, Shengxing Zhang and Jeffrey Zwiebel, as well as participants at the 2016 Stern Microstructure Conference and the 2016 EFA Doctoral Tutorial. Special thanks to Charles Zheng for reading the long proofs.

1 Introduction

Using a continuous-time model of network formation and trading in over-the-counter (OTC) markets, I show how an explicit core-periphery network arises endogenously as an equilibrium balance between trade competition and inventory efficiency. Even when agents are all ex-ante identical, a small number of them emerge as core agents, known as "dealers," who intermediate among a large number of peripheral buyside firms. The equilibrium number of dealers is determined by a key trade-off between two countervailing forces: *(i)* competition among dealers in their pricing of immediacy to buyside firms, and *(ii)* the benefits of concentrated intermediation for lowering dealer inventory risk through dealers' ability to quickly offset purchases against sales. This trade-off need not be efficient. In addition to predicting the number of dealers providing intermediation, my results point to under-provision of dealer intermediation for actively traded assets, and over-provision for infrequently traded assets.

Most OTC markets, such as those for bonds, swaps, inter-bank lending, and foreign exchange derivatives, exhibit a clear and stable core-periphery network structure.[1] Roughly the same 10 to 15 dealers, all affiliated with large banks, form the core. The vast majority of trades have one of these dealers on at least one side. For example: The largest sixteen derivatives dealers, known as the "G16,"[2] intermediate 53% of the total notional amount of interest rate swaps, 62% of credit default swaps, and 40% of foreign exchange forwards.[3] Figure 1 illustrates some examples of core-periphery networks in OTC markets.

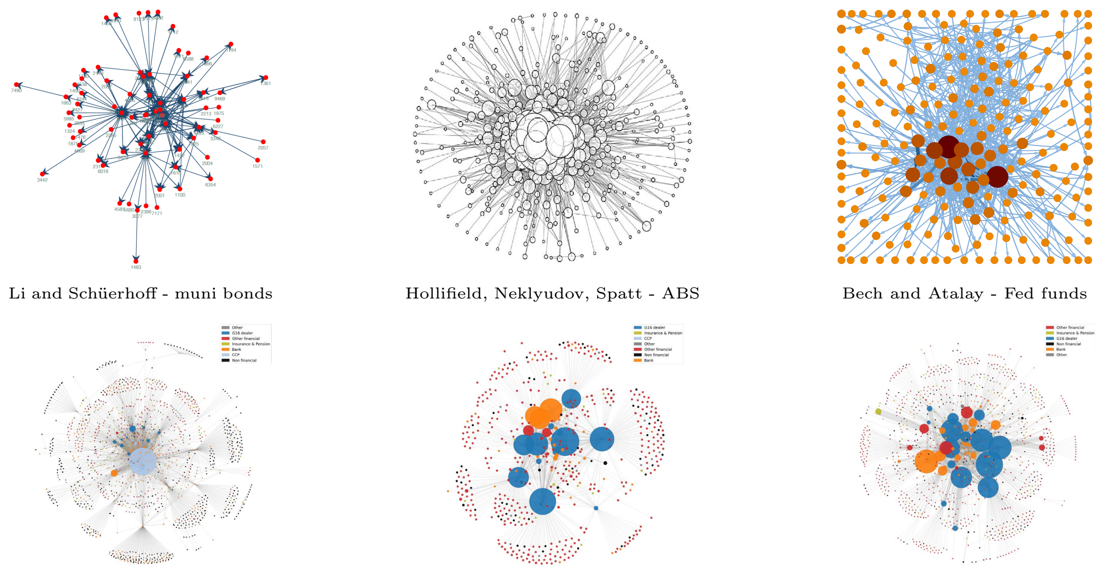
Many studies[4] have argued that recent illiquidity in bond markets has been worsened by

[1] Bech and Atalay (2010), Allen and Saunders (1986), Afonso, Kovner, and Schoar (2014) provide evidence on federal funds, Boss, Elsinger, Summer, and Thurner (2004), Chang, Lima, Guerra, and Tabak (2008), Craig and von Peter (2014), in 't Veld and van Lelyveld (2014), Blasques, Bräuning, and van Lelyveld (2015) on foreign interbank lending, Peltonen, Scheicher, and Vuillemey (2014) on credit default swaps, Di Maggio, Kermani, and Song (2015) on corporate bonds, Li and Schürhoff (2014) on municipal bonds, Hollifield, Neklyudov, and Spatt (2014) on asset-backed securities and James, Marsh, and Sarno (2012) on currencies.

[2] The G16 dealers are BoA, Barclays, BNP Paribas, Citi, Crédit Agricole, Credit Suisse, Deutsche Bank, Goldman Sachs, HSBC, JPMorgan, Morgan Stanley, Nomura, RBS, Société Générale, UBS and Wells Fargo.

[3] These statistics are computed by Abad, Aldasoro, Aymanns, D'errico, Fache, Hoffmann, Langfield, Neychev, and Roukny (2016) using EMIR data as of November 2015.

[4] Adrian, Fleming, Goldberg, Lewis, Natalucci, and Wu (2013) provide a recent discussion. Prior studies include Grossman and Miller (1988), Gromb and Vayanos (2002, 2010), Lagos, Rocheteau, and Weill (2008), Brunnermeier and Pedersen (2008), Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010), Hendershott and Seasholes (2007), Hendershott and Menkveld (2014), Rinne and Suominen (2010, 2011),



Li and Schüerhoff - muni bonds | Hollifield, Neklyudov, Spatt - ABS | Bech and Atalay - Fed funds

ESRB - Interest rate swaps | ESRB - Credit default swaps | ESRB - FX forwards

Figure 1 – Core-periphery networks in OTC markets

crisis-induced regulations (such as the Volcker Rule) and higher bank capital requirements, which have increased the cost of access to dealers' balance sheets. My results suggest that, aside from financial stability benefits (which I do not model), weighting capital requirements by asset liquidity can foster more efficient provision of dealer intermediation.

The model works as follows. A finite number of ex-ante identical agents form bilateral trading relationships in a continuous-time trading game. It is costly for agents to hold asset inventory beyond their immediate needs. Dealers arise endogenously to form the core of the market, exploiting their central position to balance inventory risk by quickly netting many purchases against many sales. Dealers compete in their pricing of immediacy to maintain long term trading relationships with peripheral buyside firms. As more dealers compete for trades, each dealer must post a narrower bid-ask spread, while requiring a higher intermediation compensation given its reduced ability to balance inventory. The equilibrium number of dealers is such that the equilibrium spread, driven by trade competition, is just enough to cover the dealer sustainable spread driven by inventory balancing. Figure 2 depicts an example equilibrium core-periphery network of 23 agents, 3 of whom emerge as dealers.

Weill (2007), Meli (2002), Adrian, Moench, and Shin (2010), Adrian, Etula, and Shin (2010).

Each dealer induces a negative externality on other dealers' inventory efficiency by reducing their order flow. This externality pushes toward over-provision of dealer intermediation, and is particularly pronounced for infrequently traded assets which have fewer opportunities for netting. For actively traded assets, however, this externality is inconsequential relative to the distortion caused by the market power of dealers over their customers. The bilateral nature of OTC trading gives dealers a temporary monopolistic position during each contact with buyside firms, causing a "holdup" distortion by which dealers extract rents that discourage some beneficial trades. For actively traded assets, the holdup effect dominates the inventory-efficiency externality, leading overall to under-provision of dealer intermediation.

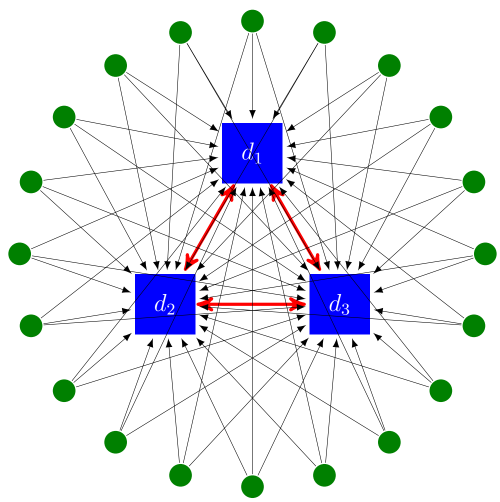


Figure 2 – An example of a core-periphery network with 3 dealers and 20 buyside firms

Partly in response to post-crisis regulation, the basic core-periphery network of some OTC markets includes additional structure in the form of trading platforms on which multiple dealers provide quotes. Multilateral trading platforms have appeared in OTC markets for foreign exchange, treasuries, some corporate bonds, and (especially through the force of recent regulation) standardized swaps. Examples of such platforms include MarketAxess and Neptune for bonds, 360T and Hotspot for currencies, and Bloomberg for swaps. This paper restricts its focus, however, to the more "classical" case of purely bilateral OTC trade.

There is a rising interest in providing theoretical foundations for the endogenous core-periphery structure of OTC markets. In prior research on this topic, the agents who form the core have some ex-ante special advantages in serving this role. Hugonnier, Lester, and Weill (2016) and Chang and Zhang (2016) derive the "coreness" of investors from their preferences for ownership of the asset. Those with average preferences act as intermediaries between high and low-value investors. The models of Neklyudov (2014) and Üslu (2016), instead, are based on exogenous heterogeneity in investors' search technologies. Moreover, prior work all leads to a continuum of "core" agents, thus ruling out realistic predictions related to the number of dealers intermediating a given market, and neglecting some strategic behavior of dealers arising from their individual impacts on the market. For example, Farboodi, Jarosch, and Shimer (2016), who allow investors to acquire superior search technologies, lead to equal equilibrium value for all core and peripheral investors. As a further distinction, in Farboodi (2015), the endogenous network structure is generated by counterparty default risk management, and not (as in my model) by trade competition and inventory risk management.

My results contribute to this literature in three ways: First, I provide a non-cooperative game-theoretic foundation for the formation of core-periphery networks in OTC markets that is motivated by inventory management and trade competition. Even when agents are all ex-ante identical, an ex-post separation of core from peripheral agents is determined solely by endogenous forces that tend to concentrate the provision of intermediation. Second, I explicitly calculate the equilibrium number of dealers as a function of market characteristics. The endogenous set of dealers has significantly higher equilibrium values than peripheral buyside firms. Finally, my model characterizes the endogenous relationships among welfare, dealer intermediation and asset trade frequency, pointing to under-provision of intermediation for actively traded assets, and over-provision for infrequently traded assets.

The paper is organized as follows. Section 2 presents the setup of the symmetric-agent model and defines the equilibrium solution concept. Section 3 shows that a core-periphery network structure emerges in equilibrium, and solves for the endogenous number of dealers as a function of market characteristics. Section 4 provides comparative statics and welfare analysis, and discusses policy implications. Section 5 offers an extension of the symmetric-agent model, in which dealers are allowed to bilaterally negotiate the terms of their trades, rather than merely offer take-it-or-leave-it quotes. Section 6 provides concluding remarks.

2 The Basic Model

Asset and preferences. I fix a probability space and the time domain $[0, \infty)$. A finite number n of ex-ante identical risk-neutral agents trade a non-divisible asset. The asset generates a sequence $(D_k)_{k \geq 1}$ of per-unit lump-sum payoffs, independent random variables with some finite mean v, at the event times of an independent Poisson process. To simplify exposition, I assume that $v = 0$, which is without loss of generality. Every agent has 0 initial endowment of the asset and incurs a quadratic cost βx^2 per unit of time when holding an asset inventory[5] of size x. That is, the agent experiences an instantaneous disutility when her inventory position deviates from a bliss point, which is normalized to 0. All agents are infinitely-lived with time preferences determined by a constant discount rate r, and can borrow and lend in a frictionless money market at the risk-free rate r.

Network formation, search and trade protocols. Each agent i is shocked by exogenously determined needs to buy or sell (equally likely) one unit of the asset at the event times of a Poisson process, independent of asset payoffs and across agents, with some mean rate 2λ. Upon receiving such a shock, there is an immediacy benefit π to agent i if the trade can be executed immediately. If it cannot, the opportunity is lost. These demand shocks can be viewed as outside customer orders, arbitrage opportunities or private hedging needs.

At any time $t \geq 0$, a given agent i can open a trading account with any other agent j, giving i the right to obtain executable price quotes from j. If i does so, then j is said to be a *quote provider* to i. Setting up a trading account is costless, but maintaining an account incurs an ongoing cost of c per unit time to agent i, which can be viewed as a monitoring or operational cost. An agent is permitted to terminate any of her accounts at any time, thus eliminating the associated maintenance costs. On the equilibrium path, these trading accounts, once set up, will be maintained forever. The option to close an account,

[5] Broker-dealers and asset-management firms have extra costs for holding inventory of illiquid risky assets. These costs may be related to regulatory capital requirements, collateral requirements, financing costs, and the expected cost of being forced to raise liquidity by quickly disposing of inventory into an illiquid market. The quadratic-holding-cost assumption is common in both static and dynamic trading models, including those of Vives (2011), Rostek and Weretka (2012) and Du and Zhu (2014).

however, plays an important equilibrium role in supporting competition as a credible threat that discourages quote providers from offering aggressively unfavorable prices.

At any time $t > 0$, agent i may search among her current quote providers. Search is cost-free, but an agent is allowed to search only a finite number of times during any finite time interval. Whenever agent i searches among her m current quote providers, there is some probability θ_m of immediate success, in which case one of these m quote providers is selected, each with equal probability $1/m$, to provide a quote. These search outcomes are independent of asset payoffs, demand shocks, and across searches. The probability $\theta_m \in (0, 1)$ of a successful search is increasing and strictly concave in $m > 0$, and $\theta_0 = 0$. Establishing more trade relationships benefits an agent's search prospects but also raises maintenance costs. Appendix B provides an example microfoundation of this search technology.

At the point of a successful search contact with some quote provider j, agent i submits a request for quote (RFQ) indicating a desired trade direction (buy or sell). Agent j then posts an executable bid or ask quote, a binding take-it-or-leave-it offer to buy or sell one unit of the asset at the respective prices. The quote is observed and executable only by agent i, and is good only when offered. The identity of agent i is not revealed to agent j at the time of the RFQ. This trade protocol is known[6] as "anonymous RFQ." Section 5 considers "name give-up RFQ," in which i "gives up" her identity to j. The restriction to a trade size of one unit is not realistic, especially for inter-dealer trading, and will be relaxed in Section 5.

When agent i accepts an ask quote of a from agent j at time t, the current inventory x_{jt} of agent j is reduced by 1 and the agreed price a is immediately transferred from i to j. Conversely, if a bid price b is accepted, x_{jt} increases by 1 and j pays[7] the amount b to i.

For technical modeling convenience, I allow a quote provider the option, whenever a quote is accepted, of not taking the trade on her own account, instead allowing the transaction with agent i to be diverted to a neutral third-party account called a "deep pocket." Agent i

[6] On Swap Execution Facilities in 2014, 31% of investors prefer trading via anonymous RFQ, and 52% prefer name give-up RFQ, according to McPartland (2014) based on survey responses.

[7] These price transfers can be immediately consumed or invested in the money market account for later consumption. Since the time discount rate r is equal to the money-market interest rate, whether to consume or save the price transfer is a matter of indifference to agent j, thus simplifying the model.

does not learn whether or not agent j invokes this deep pocket. In equilibrium, it turns out that the deep pocket is invoked only when the inventory of agent j is so large in magnitude that agent j, if she had no access to the deep pocket, would have provided a quote that would be refused by agent i, thus revealing to agent i that agent j must have an inventory that is correspondingly large in magnitude. This information would change the future search strategy of agent i in an intractable way. (Agent i would learn that seeking a future trade with agent j is relatively more likely to be a waste of time.) The existence of a deep pocket simply avoids this informational complexity. In equilibrium, as I will show, the deep-pocket account is technically feasible, in the sense that it can indeed service all trades diverted to it while maintaining a non-negative net present expected discounted value at all times.

Two tie-breaking rules are assumed: *(i)* If agent j is indifferent to using or not using her deep pocket, she does not use it; *(ii)* Whenever using her deep pocket, j always quotes a price that yields the maximum profit for her deep pocket, subject to maximizing her own continuation utility. Section 5 eliminates the deep-pocket assumption in a richer model, where the equilibrium outcome remains qualitatively intact.

Figure 3 illustrates the order of events from the perspective of a given agent i. Figure 4 shows the sequence of events that *could* happen at a given time $t > 0$.

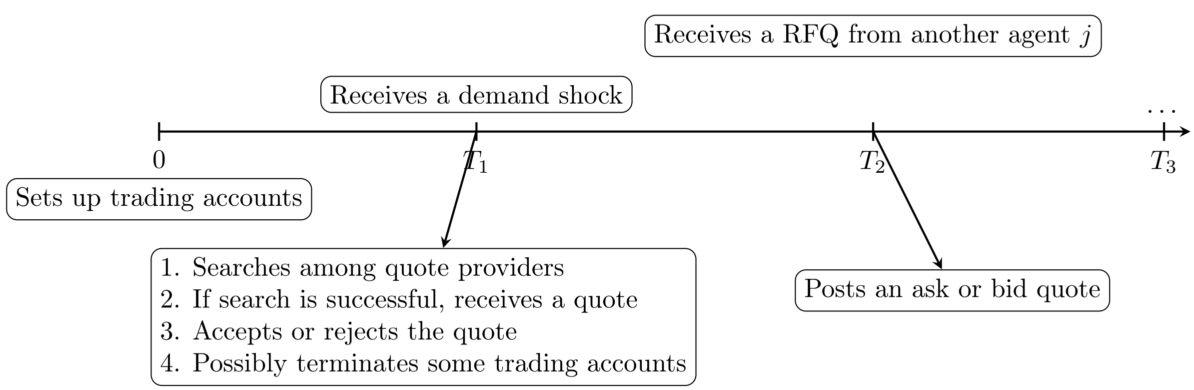


Figure 3 – Timeline of a given agent i

Information structure and solution concept. For any agent i and time $t \geq 0$, I let N_{it}^{out} be the set of her quote providers, and N_{it}^{in} be the set of agents who have i as a quote

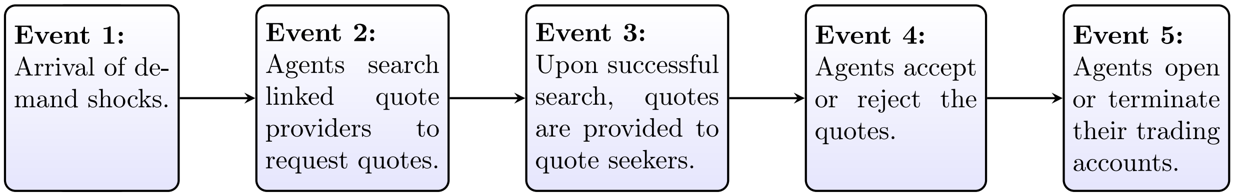


Figure 4 – Sequence of events that could occur at a given time $t > 0$

provider. The process $(N_{it})_{t\geq 0} = \left(N_{it}^{\text{out}}, N_{it}^{\text{in}}\right)_{t\geq 0}$ of potential counterparties of agent i is taken to be right continuous with left limits (RCLL).[8] I let $\mathcal{F}_{it}$ represent the information available to agent i up to but excluding time t, consisting of the sets $(N_{is})_{s<t}$ of the agent's prior counterparties, her past inventories $(x_{is})_{s<t}$, the directions $(O_{is})_{s<t}$ of her prior demand shocks (buy or sell), the requests for quote $(R_{is})_{s<t}$ from and to other agents, the quotes $(p_{is})_{s<t}$ that she has offered, the quotes $(\tilde{p}_{is})_{s<t}$ that she was offered by others, the identities $(j_{is})_{s<t}$ of the associated quote providers, and the payments $(\mathcal{P}_{is})_{s<t}$ to agent i in past trades. The payment $\mathcal{P}_{it}$ could be either the price transferred at a trade or the benefit π of fulfilling a demand shock. I let N be the set of agents. A strategy for agent i consists of

(i) A search strategy S_i that specifies, for every time $t > 0$, a search decision $S_{it} \in \{\text{Search, Do Not Search}\}$. When making this decision, agent i possesses her prior information $\mathcal{F}_{it}$, and has also observed the direction of her demand shock at time t, if there is one. Therefore, S_{it} must be measurable with respect to $\mathcal{F}_{it}^1$, the information generated[9] by $\mathcal{F}_{it}$ and O_{it}.

(ii) A quoting strategy p_i that specifies, for every time $t > 0$, the price p_{it} that i would quote upon receiving a request for quote. The quote p_{jt} is measurable with respect to $\mathcal{F}_{jt}^2$, the information generated by $\mathcal{F}_{jt}^1$ and R_{jt}. In particular, the quote is allowed to depend on whether the RFQ is to buy or sell.

[8] An RCLL function is a function defined on $\mathbb{R}^+$ that is right-continuous and has left limits everywhere. RCLL functions are standard in the study of jump processes. Please see, for example, Protter (2005). Here, the process $(N_{it})_{t\geq 0}$ is taken to be RCLL to be consistent with real-life account maintenance behavior. It also ensures that the set N_{it-} of counterparties available at time t is well defined for every $t > 0$.

[9] That is, $\mathcal{F}_{it}^1 = \sigma(\mathcal{F}_{it}, O_{it})$.

(iii) A quote acceptance strategy ρ_i for agent i specifies, for every time $t > 0$, a trade decision $\rho_{it} \in \{\text{Accept, Reject}\}$. The response ρ_{it} is measurable with respect to $\mathcal{F}^3_{it}$, the information generated by $\mathcal{F}^2_{it}$, p_{it}, $\tilde{p}_{it}$, and j_{it}.

(iv) A set $N^{\text{out}}_{it} \subseteq N \backslash \{i\}$ of quote providers to i that is measurable with respect to $\mathcal{F}^4_{it}$, the combined information of $\mathcal{F}^3_{it}$, $\mathcal{P}_{it}$, and x_{it}.

Given strategies for all agents, the continuation utility of agent i at time t is

$$U_{it} = \mathrm{E}\left(\int_t^\infty e^{-r(s-t)} \left(-\beta x^2 - \left| N^{\text{out}}_{is} \right| c \right) \, ds + \sum_{\tau_k \geq t} e^{-r(\tau_{ik}-t)} \, \mathcal{P}_{i\tau_{ik}} \, \middle| \, \mathcal{F}_{it} \right), \tag{1}$$

where $(\tau_{ik})_{k\geq 1}$ is the sequence of trade times of agent i.

In a perfect Bayesian equilibrium (PBE), each agent maximizes her continuation utility at each time, given the strategies of other agents. Appendix C provides a basic definition of PBE for continuous-time games. I focus on Markovian and stationary strategies. Formally, the Markov state variable of agent i is $Y_{it} = (x_{it}, N_{it}, O_{it}, R_{it}, \tilde{p}_{it}, j_{it})$. In a stationary equilibrium, the strategies of agent i can be written as $[f_i(Y_{it})]_{t>0}$ for some measurable function f_i.

A network G is an *equilibrium trading network* if it is supported by some stationary equilibrium σ, in that there is directed link from i to j if and only if i has a trading account with j at any time $t \geq 0$. In this case, σ is a said to be a *supporting equilibrium* for G.

3 Core-Periphery Network and Core Size

Given a set of model parameters $(n, \beta, \pi, \lambda, \theta, c, r)$, I determine all equilibrium networks, showing that they all have a flavor of "core-periphery" structure. I also provide equilibrium selection criteria that select a unique equilibrium core-periphery network.

A family of concentrated core-periphery networks.

There exists a family of equilibrium core-periphery networks of the form depicted in Figure 2. In each such network, agents are partitioned into $I \cup J = N$ with $|J| = m$ "dealers" and $|I| = n - m$ "buyside firms." Each buyside firm $i \in I$ opens a trading account with all m dealers in J. The dealers set up accounts only with each other. This network, denoted by $G(m)$, is called a *concentrated core-periphery network.* The family of equilibria is indexed by the number m of dealers, ranging from 0 to some maximally sustainable number m^* of dealers, where m^* is endogenously determined. For every $d \leq m$, I let

$$\Phi_{d,P^*(m)} = \frac{2\lambda\theta_d(\pi - P^*(m)) - dc}{r}, \qquad \text{where } P^*(m) = \pi - \frac{c}{2\lambda(\theta_m - \theta_{m-1})}. \tag{2}$$

Theorem 1. *(i) The concentrated core-periphery network $G(m)$ with m dealers is an equilibrium network if and only if $m \leq m^*$, for some maximum number m^* of dealers. (ii) In any supporting equilibrium of $G(m)$, each dealer j always posts some constant ask a_j^* and bid b_j^*, with a spread $a_j^* - b_j^* = 2P^*(m)$. The equilibrium payoff of every buyside firm is $\Phi_{m,P^*(m)}$.*

I first construct a supporting equilibrium $\sigma^*(m) = (S^*, \rho^*, p_m^*, N_m^*)$ for $G(m)$, then calculate the maximum core size m^*. In the supporting equilibrium $\sigma^*(m)$, dealers post symmetric bid-ask quotes $[-P^*(m), P^*(m)]$. In any other supporting equilibrium, dealers' bid-ask quotes do not have to be symmetric, whereas the equilibrium spread must be $2P^*(m)$.

Each agent $i \in N$ searches among her dealer couterparties when receiving a demand shock, and does not search otherwise. I denote this search strategy by S^*. When buying, i accepts any ask $a \leq \pi$. When selling, i accepts any bid $b \geq -\pi$. I denote this quote acceptance strategy by ρ^*. In equilibrium, i obtains the ask quote $P^*(m)$ when buying and the bid quote $-P^*(m)$ when selling. Thus, i earns a net profit of $\pi - P^*(m)$ for every successful execution of trade opportunity. Hence, the continuation utility of i is

$$\Phi_{m,P^*(m)} = \frac{2\lambda\theta_m(\pi - P^*(m)) - mc}{r}.$$

The numerator is the mean rate of benefit, which is the product of the demand rate 2λ, the probability θ_m of a trade, and the profit $\pi - P^*(m)$ of a successful trade, net of the account maintenance cost mc. Multiplying by $1/r$ converts this mean benefit rate to the associated lifetime present value. I now suppose that i discontinues its account with dealer d_1 at some time t. This is an off-the-equilibrium-path event that is observed only by i and d_1. Thus, all other agents cannot condition their strategies on this deviation in the continuation game. In particular, $d_2, \ldots, d_m$ continue to post the bid-ask prices $\pm P^*(m)$ to their buyside customers, including i who now only has $m-1$ dealer accounts left. The continuation utility of i thus becomes $\Phi_{m-1,P^*(m)}$. One has $\Phi_{m,P^*(m)} \geq \Phi_{m-1,P^*(m)}$, since by definition of equilibrium, m is the optimal number of dealer accounts for i. If $\Phi_{m,P^*(m)} > \Phi_{m-1,P^*(m)}$, then a given dealer would be strictly better off widening its spread by some amount $\varepsilon > 0$, knowing that it is optimal for every buyside firm to *always* maintain m dealer accounts. Therefore, the indifference condition

$$\Phi_{m,P^*(m)} = \Phi_{m-1,P^*(m)}$$

must hold for buyside firms. This indifference condition gives every buyside firm the ability to costlessly discontinue any given dealer account, making termination of a trading relationship a credible threat to the m dealers should they offer aggressively unfavorable prices. This indifference condition uniquely determines the equilibrium mid-to-bid spread $P^*(m)$ as given by (2). Figure 5 illustrates $\Phi_{d,P}$ as a function of d for $P = P^*(m)$ and $P = P^*(m+1)$.

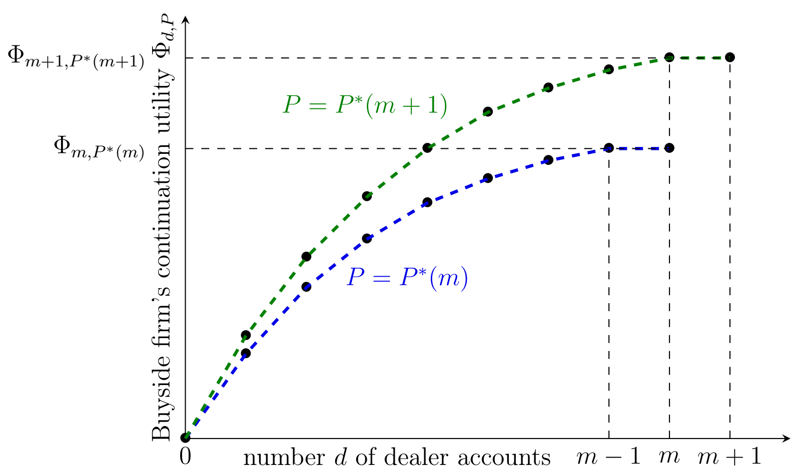


Figure 5 – Indifference condition for buyside firms – If the bid-ask spread is $P = P^*(m)$, then a buyside firm reaches its optimal contintuation value $\Phi_{d,P}$ with $d = m-1$ or m dealer accounts, while being indifferent between these two choices. If $P = P^*(m+1)$, the optimal value improves, in that $\Phi_{m+1,P^*(m+1)} > \Phi_{m,P^*(m)}$.

In $\sigma^*(m)$, agent i discontinues her trading account with a given dealer when *(a)* she has m dealer accounts in total, and *(b)* she receives an ask $a > P^*(m)$ or a bid $b < -P^*(m)$.

When these two conditions *(a)* and *(b)* simultaneously hold, agent i, after accepting the current quote if the trade is profitable (that is, if $a \leq \pi$ or $b \geq -\pi$), immediately closes her account with this dealer. In particular, any given dealer always maintains trading accounts with all other $m-1$ dealers. I denote this account maintenance strategy by N_m^*.

No account termination occurs on the equilibrium path. However, the ability of buyside firms to discontinue their dealer accounts constitutes a credible threat to dealers that discourages them from gouging. I will shortly describe the tradeoff faced by a dealer when the dealer posts quotes to a buyside firm. From expression (2) of the equilibrium spread $P^*(m)$, one obtains the following implication of trade competition by dealers.

Proposition 1. *The equilibrium mid-to-bid spread $P^*(m)$ is strictly decreasing in the number m of dealers. When there is only one dealer, the equilibrium spread $P^*(1)$ is the monopoly price, extracting all rents from buyside firms. That is, $\Phi_{1,P^*(1)} = 0$.*

As the number of dealers increases, dealers compete more intensely for trades by offering tighter bid-ask quotes. Hence, profit on each trade declines and each dealer receives a thinner order flow from buyside firms. The benefits of acting as a dealer thus decrease, limiting the equilibrium scope for dealer competition. Next, I demonstrate this intuition and determine the maximum number m^* of dealers. For this, I first describe dealer's optimization problem.

Each dealer generates the same value $\Phi_{m,P^*(m)}$ as a buyside firm for executing its exogenous trade demands. In addition, each dealer serves requests for trade from other agents. Fixing some candidate spread $P \in (0, \pi]$ and some number $k \geq 1$ of buyside customers of a given dealer $j \in J$, I consider a continuous-time control problem for j, in which price quotes are artificially restricted to the given spread P. (This restriction is later relaxed.)

Dealer's restricted-quote problem $\mathcal{P}(k, m, P)$:

(a) The state space is the set $\mathbb{Z}$ of integers, the inventory space of dealer j.

(b) The control space is $\{-P, -P_{\text{DP}}\} \times \{P, P_{\text{DP}}\}$, the set of possible bid-ask quotes that j could offer. The subscript DP denotes invoking the deep pocket. That is, dealer j

decides whether to use its deep pocket at the given ask price P upon receiving a request to buy, or at the given bid $-P$ upon receiving a request to sell.

(c) Dealer j receives RFQ from k buyside firms at the total mean rate $2k\lambda\theta_m/m$, and from the other $m-1$ dealers at the total mean rate $2\lambda\theta_{m-1}$.

(d) At each time t, dealer j maximizes, over the control space $\{-P, -P_{\mathrm{DP}}\} \times \{P, P_{\mathrm{DP}}\}$, its continuation utility U_{jt} as defined by (1).

I denote this *restricted-quote problem* by $\mathcal{P}(k, m, P)$. In the actual game, a dealer is allowed to post any quote rather than being limited to the prices $\pm P$ as in the control problem $\mathcal{P}(k, m, P)$. However, a dealer has no incentive to post any ask $a < P^*(m)$ or bid $b > -P^*(m)$, since it would otherwise cede some trading rents to its quote requester. It will later be shown that a dealer has no incentive to "gouge" by raising its ask or lowering its bid, given the fear of losing buyside customers. Therefore, the auxiliary problem $\mathcal{P}(k, m, P^*(m))$ determines an (unrestricted) optimal quoting strategy of a dealer.

I let $V_{k,m,P}$ be the value function of dealer j in the control problem $\mathcal{P}(k, m, P)$. That is, $V_{k,m,P}(x)$ is the dealer's maximum attainable continuation utility if its current inventory size is x. The Bellman principle implies that for every inventory level $x \in \mathbb{Z}$,

$$\begin{aligned} rV_{k,m,P}(x) = -\beta x^2 &+ \lambda \left(k\frac{\theta_m}{m} + \theta_{m-1}\right) [V_{k,m,P}(x+1) - V_{k,m,P}(x) + P]^+ \\ &+ \lambda \left(k\frac{\theta_m}{m} + \theta_{m-1}\right) [V_{k,m,P}(x-1) - V_{k,m,P}(x) + P]^+. \end{aligned} \tag{3}$$

The first term $-\beta x^2$ is the inventory flow cost for holding x units of the asset. The second and third terms are the expected rates of profit associated with serving requests to sell and buy, respectively, from buyside firms and the other dealers. This Hamilton-Jacobi-Bellman (HJB) equation determines a unique optimal inventory management strategy.

Proposition 2. *Given the problem $\mathcal{P}(k, m, P)$, a dealer has a unique optimal quoting strategy $[a^*(\cdot), b^*(\cdot)]$, characterized by an inventory threshold $\bar{x}_{k,m,P} \in \mathbb{Z}$ with the following property:*

- *If $x \leq -\bar{x}_{k,m,P}$, then $a^*(x) = P_{DP}$ and $b^*(x) = -P$.*

- *If* $x \geq \bar{x}_{k,m,P}$, *then* $a^*(x) = P$ *and* $b^*(x) = -P_{DP}$.

- *If* $-\bar{x}_{k,m,P} < x < \bar{x}_{k,m,P}$, *then* $a^*(x) = P$ *and* $b^*(x) = -P$.

Whenever the dealer's inventory is lower than the threshold $-\bar{x}_{k,m,P}$, it is not willing to sell more assets at the price P because the trade gain P no longer covers its indirect marginal inventory cost. Similarly, when the dealer's inventory exceeds $\bar{x}_{k,m,P}$, it is not willing to buy. If its inventory is within the range $(-\bar{x}_{k,m,P}, \bar{x}_{k,m,P})$, it has enough profit incentive to warehouse additional inventory. In equilibrium, the dealer optimally controls its inventory within the interval $[-\bar{x}_{k,m,P}, \bar{x}_{k,m,P}]$. The dealer uses its deep pocket only if its inventory is at the boundary $\bar{x}_{k,m,P}$ when receiving a request to sell, or is at the opposite boundary $-\bar{x}_{k,m,P}$ when receiving a request to buy.

Now, I consider whether dealer j has sufficient incentive to "gouge." If j decides to use its deep pocket, and if j posts some ask $a > P$, then j does not get any additional trade profit for itself (since the trade payment is received by its deep pocket), but j could lose a buyside customer. When not using its deep pocket, if dealer j "gouges" by posting some ask $a > P$, then j increases its trade profit for the current contact at the risk of losing a buyside customer and the associated future profit stream. The highest acceptable ask price being π, the one-shot benefit for the dealer of gouging is thus

$$\Pi(P) = \pi - P.$$

By symmetry, a request to sell gives the dealer the same one-shot benefit of gouging. When losing one buyside firm, the future profits forgone by j lowers its continuation value by

$$L_{k,m,P}(x) = V_{k,m,P}(x) - V_{k-1,m,P}(x).$$

The dealer has no incentive to gouge if and only if the benefit does not exceed the expected cost, in that

$$\Pi(P) \leq \mathcal{L}(k,m,P) \equiv \frac{k}{k+m-1} \min_{x \in \mathbb{Z}} L_{k,m,P}(x). \tag{4}$$

where $k/(k+m-1)$ is the probability that a contacting agent is a buyside firm (instead of another dealer), who can credibly and would discontinue its account if being gouged.

Lemma 1. *The no-gouging condition (4) is satisfied if and only if* $P \geq \underline{P}(k, m)$*, where* $\underline{P}(k, m)$ *is uniquely determined by*

$$\Pi(\underline{P}(k, m)) = \mathcal{L}(k, m, \underline{P}(k, m)).$$

Lemma 1 implies that $[-\underline{P}(k, m), \underline{P}(k, m)]$ are the tightest bid-ask quotes the dealer is willing to offer without having incentive to gouge. The mid-to-bid spread $\underline{P}(k, m)$ is called the (k, m)-*sustainable spread.* Any spread P is (k, m)-*sustainable* if $P \geq \underline{P}(k, m)$. Figure 6 illustrates the tradeoff between the gain $\Pi(P)$ and the loss $\mathcal{L}(k, m, P)$ from gouging.

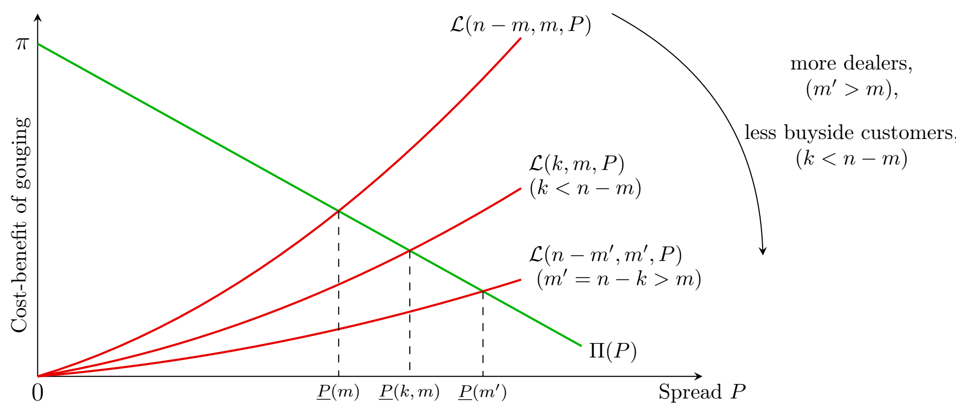


Figure 6 – The tradeoff between the one-shot benefit $\Pi(P)$ and the expected cost $\mathcal{L}(k, m, P)$ of gouging. The cost $\mathcal{L}(k, m, P)$ associated with forgone future profits is increasing in k and P and decreasing in m. Hence, the dealer-sustainable spread $\underline{P}(m)$ is increasing in m.

In the actual network trading game, a given dealer has $n - m$ buyside customers. Thus, condition (4) needs to be satisfied for $k = n - m$ for dealers to refrain from gouging. Letting $\underline{P}(m) \equiv \underline{P}(n - m, m)$, $\underline{P}(m)$ is simply called the m-*sustainable spread.*

Proposition 3. *(i) The tightest* (k, m)*-sustainable spread* $\underline{P}(k, m)$ *is strictly decreasing in the number* k *of buyside customers of a given dealer, and (ii) the tightest* m*-sustainable spread* $\underline{P}(m)$ *is strictly increasing in the number* m *of dealers.*

Intuitively, when a dealer has more buyside customers, it can offer a tighter spread thanks to its ability to efficiently balance inventory by more quickly netting purchases against sales.

A well connected dealer is in this sense a liquidity hub. When there are more dealers in the market, however, each dealer receives a thinner order flow from each given buyside firm and is not as efficient in balancing its inventory. Both factors lower a dealer's incentive to sustain a tight spread. Figure 6 provides an illustration of Proposition 3.

The equilibrium spread $P^*(m)$, as defined in (2), must be m-sustainable. Thus, $P^*(m) \geq \underline{P}(m)$. Since $P^*(m)$ is strictly decreasing in m, while $\underline{P}(m)$ is strictly increasing in m, $P^*(m) \geq \underline{P}(m)$ is equivalent to $m \leq m^*$, where m^* is the largest integer such that

$$P^*(m^*) \geq \underline{P}(m^*). \tag{5}$$

The number m^* is the *maximally sustainable core size*. Figure 7 plots both spread curves.

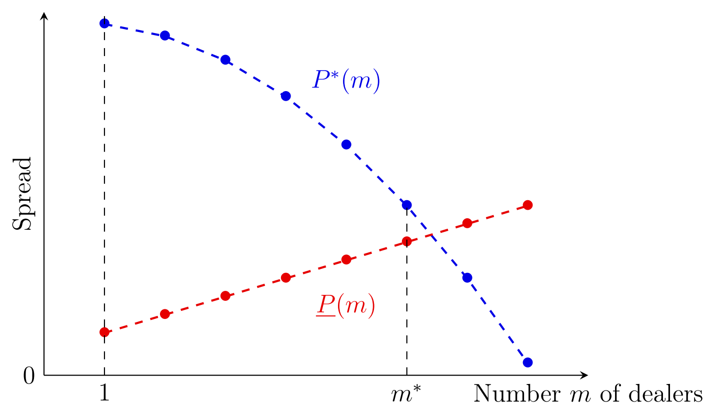


Figure 7 – The equilibrium spread $P^*(m)$, the sustainable spread $\underline{P}(m)$, and the maximum core size m^*

I let p^*_m be the optimal quoting strategy characterized by the inventory threshold $\bar{x}_{n-m,m,P^*(m)}$. If a buyside firm receives a request for quote (which is an event off the equilibrium path), it posts the ask price π and the bid price $-\pi$. I denote this quoting strategy by p^*_0.

A supporting equilibrium $\sigma^*(m)$ for $G(m)$ consists of the following strategies:

(i) Each agent follows the search strategy S^*, the quote acceptance strategy ρ^* and the account maintenance strategy N^*_m.

(ii) Dealers employ the quoting strategy p^*_m, while buyside firms employ p^*_0.

By convention, $\sigma^*(0)$ denotes the strategy profile in which no agent opens any account, and no search or trade is conducted. In the equilibrium $\sigma^*(m)$ $(m = 1, \ldots, m^*)$, dealers are deterred from gouging by the fear of losing buyside customers. Buyside firms do not terminate any account on the equilibrium path, but their ability to do so constitutes a credible threat that discourages dealers from gouging, sustaining the equilibrium.

Characterizing all equilibrium networks.

In practice, core-periphery structures in OTC markets are less "concentrated," in that a typical buyside firm is connected to some but not all dealers. Figure 8 illustrates an example of such core-periphery structure. Based on similar analysis, Theorem 2 states sufficient and necessary conditions characterizing all equilibrium networks. I show that these equilibrium networks all have a flavor of "core-periphery" structure, but are less efficient, in a sense to be specified, than the family of concentrated core-periphery networks in Theorem 1. Roughly speaking, a network is an equilibrium network if and only if *(i)* the number of dealers is relatively small, and *(ii)* every dealer has at least a minimum number of buyside customers.

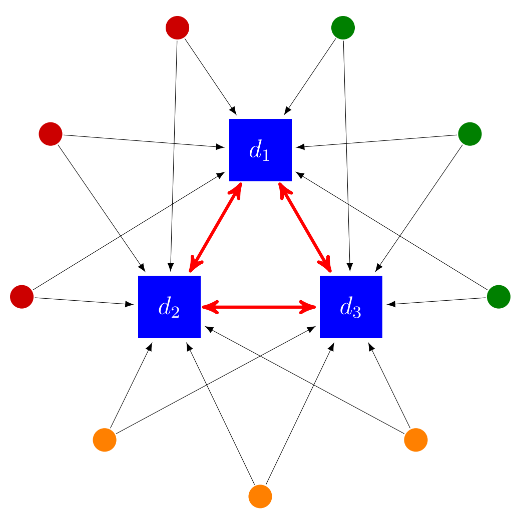


Figure 8 – Example: every buyside firm is connected to 2 of the 3 dealers

I fix a network G. Agents are partitioned into $I \cup J = N$ as follows: Agents in I, representing "buyside firms," have no incoming links; Agents in J, representing "dealers," have at least one incoming link. I denote the *maximum outdegree* of G by

$$m = \mu(G) \equiv \max_{i \in N} |N^{\text{out}}(i)|,$$

It is shown in Appendix D that every agent must have m or $m-1$ dealer accounts if G is an equilibrium network. Given a dealer $j \in J$, if k of its quote seekers have m dealer accounts each and ℓ of them have $m-1$ each, then the dealer's value function $V_{k,\ell,m,P}$ solves

$$\begin{aligned} rV_{k,\ell,m,P}(x) = -\beta x^2 &+ \lambda \left(k\,\frac{\theta_m}{m} + \ell\,\frac{\theta_{m-1}}{m-1} \right) [V_{k,\ell,m,P}(x+1) - V_{k,\ell,m,P}(x) + P]^+ \\ &+ \lambda \left(k\,\frac{\theta_m}{m} + \ell\,\frac{\theta_{m-1}}{m-1} \right) [V_{k,\ell,m,P}(x-1) - V_{k,\ell,m,P}(x) + P]^+. \end{aligned}$$

I let $\quad L_{k,\ell,m,P} = V_{k,\ell,m,P} - V_{k-1,\ell,m,P}, \qquad \mathcal{L}(k,\ell,m,P) = \dfrac{k}{k+\ell} \inf_{x \in \mathbb{Z}} L_{k,\ell,m,P}(x).$

The factor $k/(k+\ell)$ above is the probability that a given contacting agent has m dealer accounts. Only those k agents who have m dealer accounts can costlessly terminate their accounts with j. I let $\underline{P}(k,\ell,m) \in \mathbb{R}^+$ be determined by

$$\Pi(\underline{P}(k,\ell,m)) = \mathcal{L}(k,\ell,m,\underline{P}(k,\ell,m)).$$

Similar to Proposition 3, one can show that $\underline{P}(k,\ell,m)$ is strictly decreasing in k. I let $k(m,\ell)$ be the smallest integer such that

$$\underline{P}(k(m,\ell),\ell,m) \le P^*(m),$$

where $P^*(m)$ is the equilibrium spread given by (2). The dealer needs at least $k(m,\ell)$ buyside customers to sustain the equilibrium spread $P^*(m)$ without having incentive to gouge. For every given dealer $j \in J$, I denote by k_j the number of its quote seekers who each have m dealer accounts, and by ℓ_j the number of those with $m-1$ each.

Theorem 2. *A network G is an equilibrium trading network if and only if (i) every agent has m or $m-1$ dealer accounts, with $m \le m^*$, and (ii) for each dealer $j \in J$, $k_j \ge k(m,\ell_j)$. In any supporting equilibrium, every dealer posts the equilibrium mid-to-bid spread $P^*(m)$,* *and every buyside firm's equilibrium utility is* $\Phi_{m,P^*(m)}$.

Condition *(ii)* indicates that each dealer needs at least a minimum number of quote seekers to be efficient in balancing its inventory, and therefore to refrain from gouging. I will later provide an equilibrium selection criterion that selects $m = m^*$. Focusing on equilibrium networks with $m = m^*$, I derive an explicit upper bound on the total number of dealers.

Corollary 1. *If G is an equilibrium network with $m = m^*$, the number $|J|$ of dealers satisfies*

$$|J| < \frac{m^* n}{k(m^*, 0) - 1}.$$

As a numerical example, I consider a market with $n = 1000$ agents, $\beta = 0.1,\ \pi = 1,\ \lambda = 3, \theta_m = 1 - 0.8^m,\ c = 0.09$ and $r = 0.1$. The upper bound on the number of dealers is 17.

Equilibrium selection.

I propose two equilibrium selection criteria, based on inventory balancing efficiency and dealer trade competition respectively, to select the concentrated core-periphery network $G(m^*)$ with m^* dealers. The next proposition, based on inventory balancing efficiency, shows that the concentrated core-periphery networks induce higher welfare than other equilibrium networks. Given a strategy profile σ, I define welfare $U(\sigma)$ as the sum of all agents' utilities.

Proposition 4. *If σ is a supporting equilibrium for some network G that is not a concentrated core-periphery network, then there exists some supporting equilibrium σ' for a concentrated core-periphery network such that $U(\sigma') > U(\sigma)$.*

The benefit of a more concentrated network comes from higher efficiency in inventory balancing by dealers. Given an equilibrium network G in which every buyside firm has accounts with some but not all dealers, the network G' is obtained by concentrating buyside firms' accounts toward a smaller set of dealers. The networks G and G' have the same total number of trading lines, thus the same total trading volume and account maintenance cost. The same volume of trade is intermediated, however, by a smaller set of dealers in G' relative to G. Higher concentration of trades leads to more efficient netting of trades and thus a lower market-wide dealer inventory cost, which results in higher welfare in G'. The second criteria, based on trade competition, uses that the equilibria are Pareto-ranked for buyside firms.

Corollary 2. *If G and G' are two equilibrium networks, the equilibrium utility of buyside firms is strictly lower in G than in G' if $\mu(G) < \mu(G')$.*

Corollary 2 follows from Theorem 2 and $\Phi_{m,P^*(m)} < \Phi_{m',P^*(m')}$ if $m < m'$. All buyside firms prefer an equilibrium in which they are connected to more competing dealers, since the benefit associated with a tighter equilibrium spread and better trade execution outweighs additional account maintenance costs. Figure 5 provides a illustration of Corollary 2.

Corollary 2 leads to a natural equilibrium selection criterion: an equilibrium network G with maximum outdegree $\mu(G) < m^*$ can be ruled out if agents can actively coordinate the selection of dealers. For example, in the concentrated core-periphery network $G(m)$, any $m' - m$ buyside firms can credibly propose to serve as dealers, in addition to the existing m dealers, as such a proposal will result in the equilibrium network $G(m')$. In the new equilibrium, these buyside firms improve their utility by exploiting their new network positions as dealers to earn additional intermediation profits. The remaining $n - m'$ buyside firms also benefit from greater dealer competition (Corollary 2). Finally, the m existing dealers have thinner order flow and are forced to post a narrower spread $P^*(m')$. If an existing dealer refuses to lower its spread, however, it risks losing its buyside customers and ultimately its dealer position. On the other hand, the concentrated core-periphery network $G(m^*)$ with the maximum number of dealers cannot be overturned by dealer entry in this manner.

Dealer entry is a form of trade competition. The underlying assumption is that buyside firms are able to communicate their intention to serve as dealers *without* being able to commit to offer a tighter spread. The selection procedure can be formalized using the concept of *coalition-proof Nash equilibrium* introduced by Bernheim, Peleg, and Whinston (1987).

Proposition 5. *If n is sufficiently large, an equilibrium network G admits a supporting equilibrium that is a coalition-proof Nash equilibrium if and only if $\mu(G) = m^*$.*

Based on inventory balancing and trade competition respectively, Proposition 4 and Corollary 2 offer two selection criteria that select the concentrated core-periphery network $G(m^*)$.

4 Comparative Statics, Welfare and Policy Implications

To develop comparative statics on the equilibrium number of dealers, I focus attention on the concentrated core-periphery network $G(m^*)$ that is selected by Proposition 4 and Corollary 2.

The equilibrium core size m^* and the equilibrium spread $P^*(m^*)$.

The next proposition shows how the core size varies as a function of the model parameters $(n, \beta, \pi, \lambda, \theta, c, r)$. I fix all but one parameter and examine how the equilibrium number m^* of dealers, given by (5), is affected by the remaining parameter. Proofs are given in Appendix E.

Proposition 6. *(i) The core size m^* is weakly increasing in the total number n of agents, with a finite limit size m^*_∞. The limit size m^*_∞ is the largest integer m such that*

$$\frac{mr\pi}{2\lambda\theta_m + mr} < P^*(m).$$

(ii) The equilibrium number m^* of dealers is weakly increasing in the arrival rate λ of demand shocks and the total gain per trade π, and weakly decreasing in the account maintenance cost c and the inventory cost coefficient β.

(iii) The equilibrium spread $P^*(m^*)$ is weakly increasing in β, and weakly decreasing in the total number n of agents.

Part *(i)* of Proposition 6 has a simple intuitive proof, as follows. As the total number n of agents increases, each dealer becomes more efficient in balancing inventory, thus can sustain a tighter spread $\underline{P}(n-m, m)$ (Proposition 3). The equilibrium spread $P^*(m)$, however, does not depend on n. The core size m^* is thus weakly increasing in n, as shown by Figure 9.

Part *(ii)* implies that even for an "infinite" set of investors, one should anticipate only a finite number m^*_∞ of dealers. To provide a numerical example of the number m^*_∞ of dealers in a large market, I let $\pi = 1$, $\lambda = 3, \theta_m = 1 - 0.8^m$, $c = 0.09$, and $r = 0.1$. Then $m^*_\infty = 3$, and the equilibrium spread in the large market is $P^*(m^*_\infty) \simeq 0.1$.

As π increases, dealers extract a higher rent per trade (reflected by a wider equilibrium spread $P^*(m)$), but also have a stronger incentive to gouge (wider sustainable spread $\underline{P}(m)$).

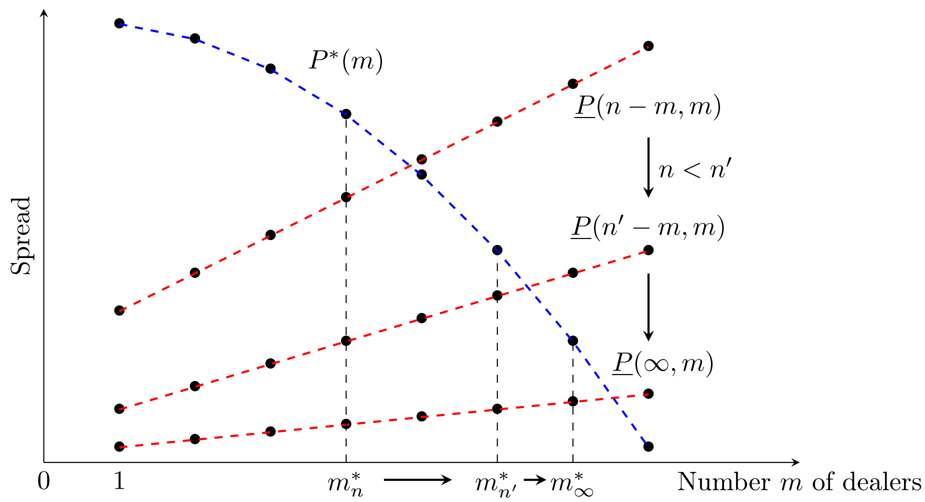


Figure 9 – The core size m^* is weakly increasing in the total number n of agents.

It is shown, in Appendix E, that the equilibrium spread $P^*(m)$ increases more than the sustainable spread $\underline{P}(m)$. The core size m^* is thus weakly increasing in π.

The parameter λ measures the liquidity demand of each agent. With a higher liquidity demand, it is natural that more dealers emerge to facilitate the intermediation of the asset, leading to a lower market concentration. This prediction of a negative relationship between liquidity demand and market concentration is consistent with empirical evidences from OTC markets. Using data on the German Bund market, de Roure and Wang (2016) show that higher trade frequency leads to lower Herfindahl index. In the foreign exchange derivatives market, the Herfindahl index ranking is, from low to high, USD, EUR, GBP, JPY, CHF, CAD and SEK. The order of the outstanding notional amounts of these currencies is almost reversed(with the exception of JPY and GBP, which are close in both measures). Across asset classes, the Herfindahl index is lowest in the interest rate derivatives market, followed by the credit derivatives market and finally the equity derivatives market. As a time-series example, Cetorelli, Hirtle, Morgan, Peristiani, and Santos (2007) document a substantial decline in the market concentration of the credit derivatives market during 2000-04, as "financial institutions have rushed to take part in this exploding market." Figures 10 to 12 and Table 1 illustrate these four examples.

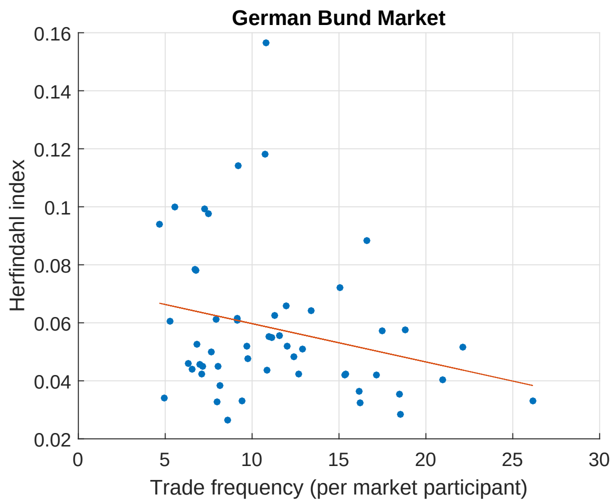


Figure 10 – Source: de Roure and Wang (2016)

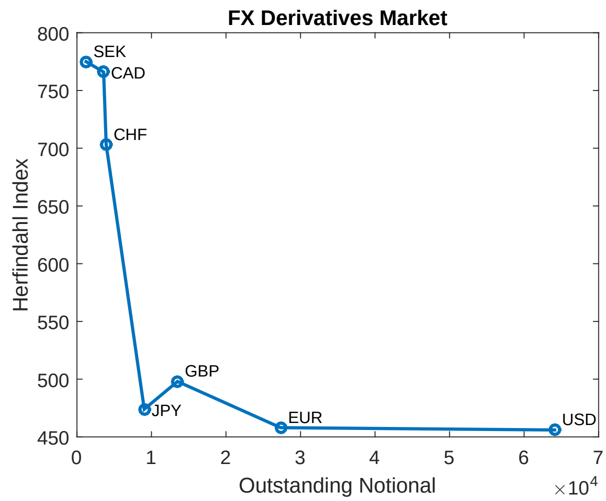


Figure 11 – Source: Semiannual Statistics (BIS)

	Four firms		Eight firms		HHI
	Notional	Percent	Notional	Percent	
Interest rate	173.5	40.0	272.9	62.9	629.4
Credit	10.7	40.8	18.4	69.9	738.5
Equity	2.7	43.0	4.5	70.8	747.9
Total	184.6	39.5	293.2	62.8	630.1

Table 1 – Souce: ISDA Market Survey, Mid-Year 2010, by Mengle (2010)

TABLE 2
Concentration Trends in Interest Rate and Foreign Exchange Over-the-Counter Derivatives Markets

Market	Average HHI	Growth in HHI (Percent)
Panel A: Global concentration: BIS surveys, 1998-2004		
U.S. interest rate derivatives		
Forward rate agreements	843	4.64
Interest rate swaps	591	8.20
Options	908	0.75
Foreign exchange derivatives		
Forwards and swaps	420	5.30
Options	544	2.31
Panel B: Federal Reserve Bank of New York Estimates of Concentration: 2000-04		
Credit derivatives	825	-14.04

Figure 12 – Source: Cetorelli, Hirtle, Morgan, Peristiani, and Santos (2007)

If dealers become less risk tolerant (higher β), they find it more costly to warehouse inventory risk. As a result, dealers need to be compensated with a wider spread $\underline{P}(m)$ to continue making markets. On the other hand, the equilibrium spread $P^*(m)$ is not affected by dealer risk tolerance, as it is determined by an indifference condition for buyside firms. Therefore, the market can only support a smaller core with agents of reduced risk tolerance.

A current hotly debated issue is bond market illiquidity. The world's biggest banks are shrinking their bond trading activities to comply with post-crisis regulations such as the Volcker rule and higher capital requirements. These restrictions have curbed the ability of banks to build inventory or warehouse risk. On Friday, October 23, 2015, Credit Suisse exited its role as a primary dealer across Europe's bond markets, the latest signal that banks are scaling back bond trading activities. Other markets, such as corporate bond and currency, are slowly experiencing structural changes due to significant dealer disintermediation. Intermediation in these markets is increasingly agency-based, and many investors report that post-crisis regulatory reforms have reduced market liquidity.

Dealer inventory levels and turnover, market-wide dealer inventory cost.

Given the model parameters $(n, \beta, \pi, \lambda, \theta, c, r)$, the total arrival rate of demand shocks is $2n\lambda$. In the concentrated core-periphery network $G(m)$ ($m \leq m^*$), I examine how the dealer inventory dynamic depends on n and λ. I let $\bar{x}$ denote the inventory threshold $\bar{x}_{n-m,m,P^*(m)}$.

Proposition 7. *The inventory threshold $\bar{x}$ is weakly decreasing in β and weakly increasing in $n\lambda$. As $n\lambda$ goes to infinity, $\bar{x}$ goes to infinity at the rate $(n\lambda)^{1/3}$, and the mixing time*[10] *of dealer inventory process goes to 0 at the rate $(n\lambda)^{-1/3}$.*

Fixing the level of dealer competition, as it becomes more costly to warehouse inventory, dealers optimally reduce their inventory size. When the asset is more liquid (either because of a larger rate λ or because of a larger number n of market participants), dealers expand their inventory size to take advantage of the increased order flow from quote seekers.

[10] The mixing time of a Markov process $(X_t)_{t\geq 0}$ is defined as $t_{\text{mix}} = \inf\{t : d(t) \leq 1/4\}$, where $d(t) = \sup_{x_0} ||X_t, \mu||_{\text{TV}}$ is the total variation distance between X_t and the stationary distribution μ of the Markov process (X_t). Levin, Peres, and Wilmer (2009) provide background on Markov chains and mixing times.

Roughly speaking, the mixing time of dealer inventory process is the expected time it takes for a dealer to rabalance its inventory. In a liquid asset market, dealer inventory has quick turnover and exhibits fast mixing. The positive relationship between asset liquidity and the rate of dealer inventory rebalancing, as predicted by the model, is consistent with prior empirical studies. Using data on the actual daily U.S.-dollar inventory held by a major dealer, Duffie (2012) estimates that the "expected half-life" of inventory imbalances is approximately 3 days for the common shares of Apple, versus two weeks for a particular investment-grade corporate bond. The data also reveal substantial cross-sectional heterogeneity across individual equities handled by the same market maker, with the expected half-life of inventory imbalances being the highest for (least liquid) stocks with the highest-bid-ask spreads and the lowest trading volume. Figure 13 illustrates the two inventory processes of the Dealer.

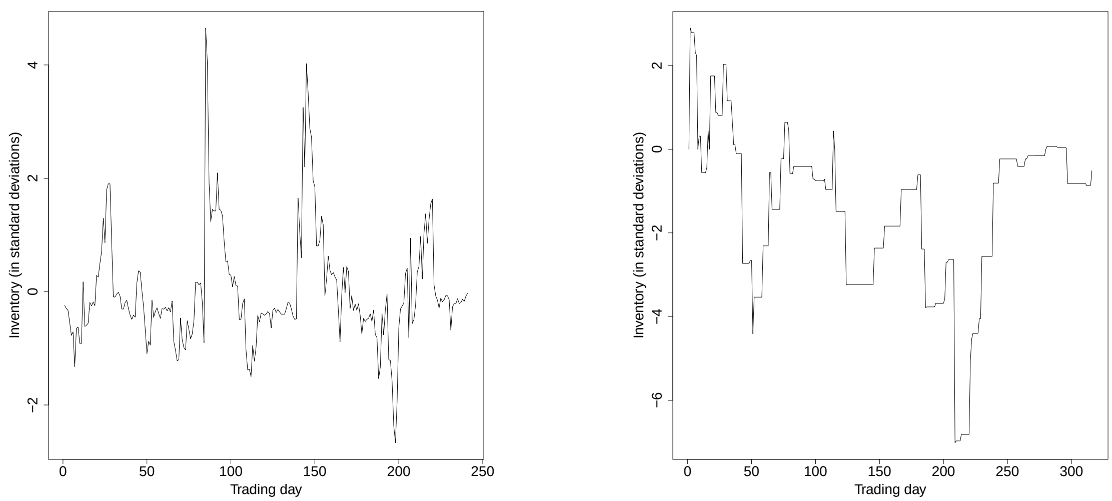


Figure 13 – Inventory processes of a major US dealer - Source: Duffie (2012)

Next, I examine properties of dealer inventory cost. The equilibrium utility of a dealer can be decomposed into inventory cost and profits from trading with quote seekers:

$$V_{n-m,m,P^*(m)}(0) = -C(n,\lambda,m) + 2\lambda\left((n-m)\frac{\theta_m}{m} + \theta_{m-1}\right)\frac{P^*(m)}{r}.$$

Proposition 8. *(i) The present value $C(n,\lambda,m)$ of individual dealer inventory cost is strictly*

concave in n and λ. As $n\lambda$ goes to infinity, $C(n, \lambda, m)$ increases to infinity at the rate $(n\lambda)^{2/3}$.
(ii) The market-wide total cost $mC(n, m, \lambda)$ of dealer inventory is strictly increasing in m.

The inventory holding cost βx^2 is quadratic in inventory size, whereas the present value $C(n, \lambda, m)$ of individual dealer inventory cost grows sublinearly with n and λ. This captures the netting benefit, in the sense that a given dealer is more efficient in balancing its inventory when receiving thicker order flow, and the associated netting effect more than offsets the convexity of the inventory cost function.

Property *(iii)* follows from the decreasing returns to scale of the individual inventory cost function $C(n, m, \lambda)$ and Jensen's inequality. It implies that in order to minimize the market-wide dealer inventory cost, it is better to concentrate the provision of intermediation at a smaller set of dealers in order to maximize the netting efficiency.

Inventory-efficiency externality and holdup distortion.

In OTC markets, it is extremely rare for regulators to directly intervene in asset allocation. However, regulators may impose transaction tax, capital requirements or some price rule to induce a different equilibrium outcome, in which decisions related to trading and link formation are still left to market participants. Subject to equilibrium selection, the concentrated core-periphery network $G(m^*)$ emerges as the unique equilibrium network in the model, where the core size m^* is endogenously determined as a function of model parameters. Therefore, feasible regulations amount to induce a different endogenous core size. Regulators thus face a one-dimensional problem, in which they choose the optimal number of dealers intermediating a given market. From a welfare viewpoint, the next result points to under-provision of dealer intermediation for liquid assets, and over-provision for illiquid assets. I discuss the effects of three regulation policies – a "soft" stub-quote rule, a transaction tax, and capital requirements – in inducing a more efficient level of dealer intermediation.

For any given integer m (possibly greater than m^*), I let $U_m = U(\sigma^*(m))$ denote the welfare induced by the strategy profile $\sigma^*(m)$, and $\overline{m}$ be the largest integer such that

$$P^*(\overline{m}) > 0.$$

I let m^{**} be the socially efficient number of dealers, in that

$$m^{**} = \underset{m \leq \overline{m}}{\operatorname{argmax}} \; U_m$$

Proposition 9. *(i) If $n\lambda$ is sufficiently large, one has $m^{**} \geq m^*$ and under-provision of dealer intermediation. (ii) Under certain parameter conditions that reduce $n\lambda$, one has $m^{**} < m^*$ and over-provision of dealer intermediation.*

These inefficiencies result from two sources. First, each dealer induces a negative externality on other dealers' inventory efficiency by reducing their order flow, as shown by Proposition 8. This inventory-efficiency externality pushes toward over-provision of dealer intermediation. Second, dealers' monopolistic position in each contact with buyside firms give them a private incentive to gouge their customers. Even though dealers do not gouge in equilibrium, their incentive to gouge induces a holdup distortion, by which dealers extract rents that discourage buyside firms from seeking socially beneficial trades. This holdup effect pushes toward under-provision of intermediation. For an actively traded asset, the total inventory cost is inconsequential relative to the welfare (Proposition 8). Therefore, the inventory-efficiency externality is dominated by the holdup distortion, leading overall to an under-provision of intermediation. For an infrequently traded asset, however, the inventory cost is large relative to the welfare. The inventory-efficiency externality could outweigh the holdup distortion, resulting overall in an over-provision of intermediation. Figure 14 shows numerically the welfare U_m as a function of m for a liquid ($\lambda = 120$, or 20 trade demands per agent per month) and an illiquid asset ($\lambda = 12$, or 2/agent/month), respectively.

To improve market efficiency, under-intermediation can be mitigated by regulations that aim to discourage dealers from gouging. Such regulations can be, for example, a "soft" stub-quote rule that imposes a penalty should a dealer widen its spread relative to the market-prevailing level. Every dealer would internalize the penalty cost C_{penalty} into its loss $\mathcal{L}(n - m, m, P)$ from gouging and can thus sustain a tighter spread $\underline{\underline{P}}(m)$. Such a penalty cost on dealers - never triggered in equilibrium - encourages dealer entry. The penalty creates room for greater dealer competition by improving dealer's commitment power. By choosing

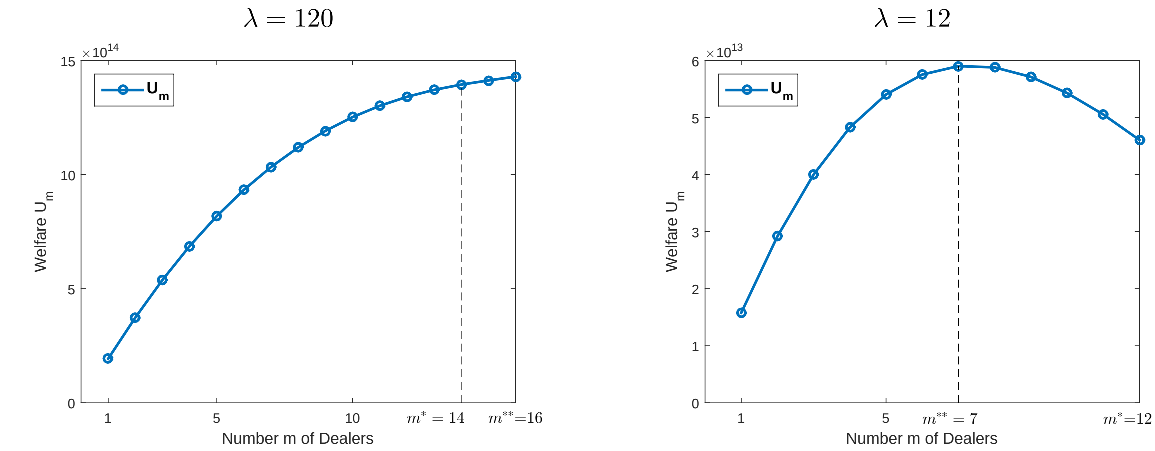


Figure 14 – The welfare U_m, where the number of dealers is $m = 1, 2, \ldots, \overline{m}$

an appropriate penalty cost C_{penalty}, regulators can achieve the socially optimal level of intermediation provision. Figure 15 illustrates the effect of such a penalty cost. Sometimes, such rules are proposed by self-regulatory organizations (SRO), such as FINRA. Broker-dealers have an incentive to join such an SRO, as the improvement of their commitment power via self regulation gives them an advantage over their non-member competitors.

To reduce dealer intermediation, regulators can impose a transaction tax on dealers, which would widen their sustainable spread, reducing the endogenous core size m^*.

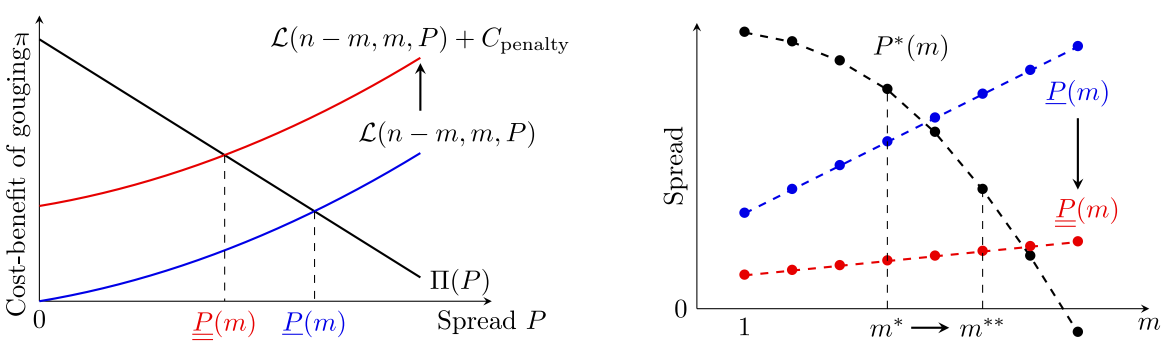


Figure 15 – Introducing a penalty cost C_{penalty} on gouging increases dealer competition.

Post-crisis regulations such as the Volcker rule and capital requirements have been implemented to limit dealer risk appetite. My results suggest that, aside from financial stability benefits (which I do not model), weighting balance-sheet regulations by asset liquidity can foster more efficient provision of dealer intermediation. To improve welfare, regulators should encourage dealer intermediation for liquid assets such as the Treasuries, and reduce intermediation for illiquid assets. The current regulatory capital requirements adopted by Basel III uses risk-weighted assets as the denominator of the capital ratio of a bank. Similarly, the Basel III Net Stable Funding Ratio (NSFR) and Supplementary Leverage Ratio (SLR)[11] treats high quality liquid assets (HQLA) equally as non-HQLA. These approaches can be improved by adding a liquidity component into the weight calculation, putting lower weights on more liquid assets in order to encourage dealer intermediation for these assets. This implication on balance-sheet regulations – based on intermediation efficiency – is in line with the primary objective of regulators in promoting financial stability, which I do not model.

Recently, more non-bank firms such as fund managers have begun to act as liquidity providers. However, many question whether these firms can substitute for dealers by taking an effective role of market makers. This paper highlights the importance of having a large customer base for a market maker to efficiently balance inventory. Being in a central network position is essential for enabling a financial institution to "lean against the wind" – that is, to provide liquidity during financial disruptions. Buyside firms are not naturally liquidity hubs. Without the same number of trading lines and global customer base that traditional dealers have, these firms may be unable or unwilling to absorb external selling pressure in a selloff. It is worrisome that the liquidity provided by non-bank firms may be "illusory," in that liquidity may vanish when it is most needed. This paper does not cover this topic.

5 Inter-Dealer Trading through Nash Bargaining

In practice, dealers trade with each other – usually in large quantities – to share their inventory risk accumulated from trading with buyside firms. Rather than a request for quote by one counterparty and a take-it-or-leave-it offer by the other, two dealers usually

[11]The NSFR, to be implemented in 2018, requires banks to maintain sufficient available stable funding (ASF) relative to the amount of required stable funding (RSF). The SLR, also to be implemented in 2018, requires U.S. globally systemically important bank holding companies to have capital equal to or greater than 5% of their total assets, regardless of the risk and liquidity composition of the assets.

negotiate both the trading quantity and price with roughly equal bargaining power. In this section, I present a more realistic variant of the symmetric-agent model, in which dealers conduct Nash Bargaining when trading with each other. The model no longer preserves agent symmetry, as a subset of agents (dealers) have access to an interdealer market with a trading protocol featuring Nash bargaining that is not available to other agents (buyside firms). In practice, dealers resist the participation of buyside firms in the interdealer segment and accuse buyside firms of taking liquidity without exposing themselves to the risks of providing liquidity. Others criticize dealers for trying to prevent competition that would compress bid-ask spreads in the market.[12] The equilibrium of the symmetric-agent model remains to be an approximate equilibrium, despite the addition of an interdealer market.

The structure of the trading game is similar to that of the symmetric-agent model of Section 2. A non-divisible asset with 0 expected payoffs is traded by n agents. Every agent has 0 initial endowment of the asset, and is subject to the quadratic inventory holding cost βx^2. The time discount rate is r. At any time $t \geq 0$, agents can open new and terminate existing trading accounts. Maintaining an account costs c per unit of time.

All assumptions above are identical to the symmetric-agent model. Next, I distinguish dealers from buyside firms and introduce an interdealer market. Agents are partitioned into $I \cup J = N$ with $|J| = m$ dealers and $|I| = n - m$ buyside firms. Every buyside firm in I has an exogenously determined desire to buy of sell (equally likely) one unit of the asset at mean rate 2λ, and receives a fixed benefit π for each immediate execution of such trade. Dealers in J do not receive demand shocks. Dealers give each other Nash bargaining rights. Specifically, if a pair of dealers $j_1, j_2 \in J$ have trading accounts with each other, they bilaterally negotiate the quantity and price according to Nash bargaining when they trade. The buyside firms have no access to this interdealer market. That is, every buyside firm only trades via RFQ. The RFQ protocol, however, does not need to be anonymous. Instead, I assume name give-up RFQ, more common in OTC markets, in which the quote requester "gives up" her identity to the quote provider. For more realism, I also eliminate the deep pocket assumption. That is, agents have no access to deep pockets. Encounters between

[12] Some recent electronic facilities such as SEF blur the exclusivity of the interdealer market.

pairs of dealers are based on independent random matching, with pair-wise meeting intensity ξ. The search technology of buyside firms remains the same as in the symmetric-agent model. That is, if a buyside firm searches among its d quote providers, there is some probability θ_d of immediate success, where $\theta_d \in (0, 1)$ is increasing and strictly concave in $d > 0$, and $\theta(0) = 0$. The information structure departs from that of the symmetric-agent model in two aspects: *(i)* The model assumes the name-give-up RFQ protocol, as mentioned above. *(ii)* When two dealers meet, they observe all dealers' inventories, which implies complete information for interdealer bargaining. The model avoids bargaining with incomplete information and maintains modeling focus on network formation and trading.

To summarize, this model is different from the symmetric-agent model in three aspects: *(i)* the introduction of an interdealer market, *(ii)* the replacement of anonymous RFQ by name give-up RFQ, and *(iii)* the elimination of the deep pocket assumption. I focus on large markets and use an approximate equilibrium concept when solving this model. This restriction is not necessary for solving the symmetric-agent model. In a perfect ε-equilibrium,[13] each agent's continuation utility at each information set at each time is within ε of her maximum attainable continuation utility, given the strategies of other agents.

Since deep pockets are no longer available, I let $\hat{p}^*_m$ denote the quoting strategy that is obtained from p^*_m by replacing deep-pocket quotes with $b = -\infty$ and $a = \infty$, prices which signal that the quote provider has no intention to buy or sell, respectively. I let $\hat{\sigma}^*(m)$ denote the strategy profile that is obtained from $\sigma^*(m)$ by replacing dealers' quoting strategy p^*_m with $\hat{p}^*_m$. The next theorem shows that, when the market is sufficiently large, introducing an interdealer market does not qualitatively affect the equilibrium outcome of the symmetric-agent model.

Theorem 3. *For a given set of model parameters $(n, \beta, \pi, \lambda, \xi, \theta, c, r)$, I let m^* be defined as in (5). There is some constant n_0, such that if $n > n_0$, then $\hat{\sigma}^*(m)$ is a perfect ε-equilibrium for every $m \leq m^*$, supporting the concentrated core-periphery network $G(m)$.*

With an interdealer market, I provide a testable prediction about interdealer volume.

[13] Mailath, Postlewaite, and Samuelson (2005) provide game theoretical background for this concept.

Proposition 10. *As $n\lambda \to \infty$, the fraction of interdealer volume is on the order of* $(n\lambda)^{-2/3}$.

For more actively traded assets, interdealer trade accounts a small fraction of total trade volume. Intuitively, the volume of an interdealer trade is large only when dealer inventories are far away from their long-run averages and need to be quickly rebalanced. When the total demand for trade is high, dealers can efficiently balance their inventories with customer orders. Therefore, dealers are less reliant on each other to lay off their inventory risk.

Using TRACE transaction data for U.S. corporate bonds between 2005-2014, I estimate the relationship between the fraction of interdealer volume and annual trade volume across all 61,823 bonds. Proposition 10 predicts that the logarithms of these two variables are linearly related, with a negative slope. Consistent with this prediction, the data shows that a 10% increase in total volume is associated with a 1% decrease in the fraction of interdealer volume. The t-statistic is -7, with standard errors clustered at the company level. This result illustrates the role of inventory efficiency in dealers' ability to provide intermediation.

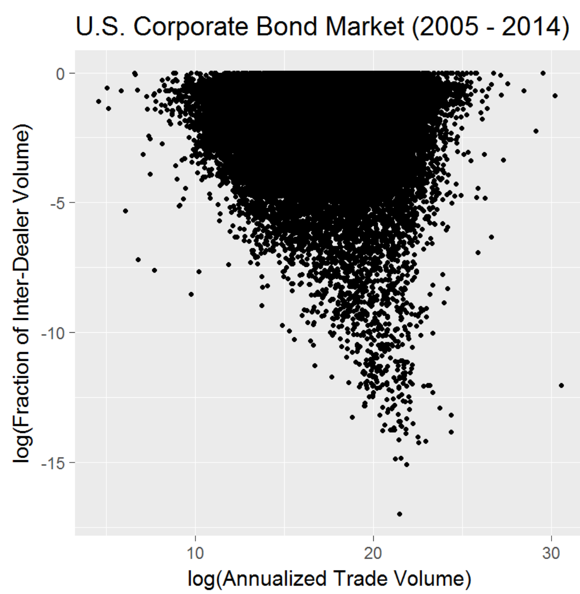


Figure 16 – Interdealer Trading in the U.S. Corporate Bond Market

To examine asset intermediation efficiency, I let $\widehat{U}_m = U(\hat{\sigma}^*(m))$ be the welfare of $\hat{\sigma}^*(m)$.

Proposition 11. *When n is sufficiently large, the welfare $\widehat{U}_m$ is strictly increasing in the number m of dealers, for $m \leq \overline{m}$. There is under-provision of dealer intermediation.*

6 Concluding Remarks

Extensive empirical work has shown that core-periphery networks dominate conventional OTC markets. However, few theoretical foundations have been provided. Existing literature has a continuum of dealers, and exploits some ex-ante heterogeneity of agents to explain the ex-post differentiation in their "network" positions. This paper is original in its ability to *(i)* provide a separation of core from peripheral agents solely based on trade competition and inventory balancing – two endogenous forces that tend to concentrate the provision of intermediation, and to *(ii)* explicitly determine the equilibrium number of dealers as a trade-off between these two forces. Although financial institutions are heterogeneous in real OTC markets, the core-periphery separation obtained in this paper highlights the importance of these two economic forces in determining market structure.

From a welfare viewpoint, the model identifies two sources of externalities: (1) dealers' private incentive to gouge, and (2) the negative externality of each individual dealer on the market-wide netting efficiency. The first pricing externatlity dominates for a liquid asset and leads to insufficient dealer intermediation. Regulators or an SFO can implement a soft stub-quote rule to deter dealers from gouging. Such a price rule improves dealers' commitment power, and therefore creates room for greater dealer competition. The second inventory externality is more pronounced for an illiquid asset, and results in over-provision of dealer intermediation. These welfare results suggest balance-sheet regulations that treat assets differently according to their liquidity demand through, for example, the introduction of a "liquidity weight," in addition to the currently adopted "risk weight."

One useful direction of future research is to introduce agent heterogeneity in order to study the relationship between dealer centrality and the pricing of immediacy. Recent empirical work suggests that the price-centrality relationship changes across different markets. In the municipal bond market, central dealers earn higher markups compared with less central dealers.[14] The opposite is true in the market for asset-backed securities.[15]

[14]Li and Schürhoff (2014) provide evidence from the municipal bond market.

[15]Hollifield, Neklyudov, and Spatt (2014) provide evidence from the market of asset-backed securities.

Appendices

A A Symmetric Tri-Diagonal Matrix

This appendix establishes some properties for the inverse of a symmetric tri-diagonal matrix. I let n be a strictly positive integer. A vector ψ of length n is said to be *U-shaped* if

$$\psi_i = \psi_{n+1-i}, \qquad \forall\, 1 \le i \le n, \qquad \text{and} \quad \psi_1 > \psi_2 > \cdots > \psi_m, \text{ where } m = \left\lfloor \frac{n+1}{2} \right\rfloor.$$

Given two vectors ψ and φ of the same length, I write $\psi < \varphi$ if ψ is strictly less than φ entry-wise. Given a constant $\zeta > 1$, I let A be the following tri-diagonal matrix of size $n \times n$:

$$A = \begin{pmatrix} \zeta - \frac{1}{2} & -\frac{1}{2} & & & & \\ -\frac{1}{2} & \zeta & -\frac{1}{2} & & & \\ & -\frac{1}{2} & \zeta & -\frac{1}{2} & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\frac{1}{2} & \zeta & -\frac{1}{2} \\ & & & & -\frac{1}{2} & \zeta - \frac{1}{2} \end{pmatrix} \tag{6}$$

Lemma 2. *The matrix A is invertible. Its inverse $M \equiv A^{-1}$ satisfies the following properties:*

(i) The matrix M is symmetric, with strictly positive entries.

(ii) For every $1 \le i \le n$, $M_{i,1} < M_{i,2} < \cdots < M_{i,i}$ and $M_{i,i} > M_{i,i+1} > \cdots > M_{i,n}$.

(iii) For every $i \ne j$, $M_{i,j-1} + M_{i,j+1} > 2M_{i,j}$.

(iv) For every i, j, $M_{i,j} = M_{n+1-i,n+1-j}$.

(v) If a vector ψ is U-shaped, then $M\psi$ is U-shaped and $(M\psi)_n \le \psi_n/(\zeta - 1)$.

(vi) Letting $m = \lfloor (n+1)/2 \rfloor$, then $M_{m,1} - M_{n+1-m,1} < 2$.

(vii) If $n \to \infty$ and $\zeta \to 1$ with $(\zeta - 1)n \to 0$, letting $\operatorname{range} M = \max_{(i,j)} M_{ij} - \min_{(i,j)} M_{ij}$,

$$(\zeta - 1) \operatorname{range} M \sim (\zeta - 1)n.$$

Proof. Properties (i). Since $\zeta > 1$, the matrix A is diagonally dominant thus invertible. Its inverse M is symmetric since A is. The matrix A can be written as $A = \zeta I - B/2$, where

$$B = \begin{pmatrix} 1 & 1 & & & \\ 1 & & 1 & & \\ & \ddots & & \ddots & \\ & & 1 & & 1 \\ & & & 1 & 1 \end{pmatrix}$$

The sup-norm of the matrix B is $||B||_\infty = 2$. All entries of M are strictly positive, since

$$M = A^{-1} = \zeta^{-1} \left(I - \frac{B}{2\zeta} \right)^{-1} = \zeta^{-1} \left[I + \frac{B}{2\zeta} + \left(\frac{B}{2\zeta} \right)^2 + \dots \right].$$

Properties (ii) and (iii). One has $MB/2 = \zeta M - I$. Then for every $i > 1$,

$$\frac{M_{i,1} + M_{i,2}}{2} = \zeta M_{i,1} > M_{i,1} \implies M_{i,1} < M_{i,2}.$$

I suppose $M_{i,j-1} < M_{i,j}$ for some $j \in (1, i)$, then

$$\frac{M_{i,j-1} + M_{i,j+1}}{2} = \zeta M_{i,j} > M_{i,j} \implies M_{i,j} < M_{i,j+1}.$$

By induction, one has $M_{i,j} < M_{i,j+1}$ if $j < i$. Similarly, one has $M_{i,j} < M_{i,j-1}$ if $j > i$.

Property (iv). It is clear that $B_{i,j} = B_{n+1-i,n+1-j}$ for every i, j. If $B^\ell_{i,j} = B^\ell_{n+1-i,n+1-j}$, then

$$B^{\ell+1}_{i,j} = \sum_k B_{i,k} \left(B^\ell \right)_{k,j} = \sum_k B_{n+1-i,n+1-k} \left(B^\ell \right)_{n+1-k,n+1-j} = B^{\ell+1}_{n+1-i,n+1-j}.$$

Therefore, $M_{i,j} = M_{n+1-i,n+1-j}$ for every i, j.

Property (v). Given a U-shaped vector ψ, then for every i,

$$(M\psi)_i = \sum_j M_{i,j} \psi_j = \sum_j M_{n+1-i,n+1-j} \psi_{n+1-j} = (M\psi)_{n+1-i}.$$

For every $0 \leq k \leq m$, I let

$$w(k) = \Bigg(\underbrace{1,\ldots,1}_{k\ 1's},\ \underbrace{0,\ldots,0}_{(n-2k)\ 0's},\ \underbrace{1,\ldots,1}_{k\ 1's}\Bigg)^\top.$$

Any U-shaped vector ψ can be written as a linear combination of the vectors $w(k)$ with strictly positive weights. Thus, to show that $M\psi$ is U-shaped for any U-shaped vectors ψ, it is sufficient to show that $Mw(k)$ is U-shaped for every $0 \leq k \leq m$. For every $i \in [k, m)$,

$$\begin{aligned}[Mw(k)]_{i+1} &= \sum_{j\leq k} M_{i+1,j} + \sum_{j>n-k} M_{i+1,j} \\ &= \sum_{j\leq k} (M_{j,i+1} + M_{j,n-i}) < \sum_{j\leq k} (M_{j,i} + M_{j,n-i+1}) = [Mw(k)]_i.\end{aligned}$$

I let $e = (1,\ldots,1)^\top$. Then $Ae = (\zeta - 1)e$ and thus $Me = e/(\zeta - 1)$. Then for every $i < k$,

$$[Mw(k)]_{i+1} = \frac{1}{\zeta - 1} - \sum_{k<j\leq n-k} M_{j,i+1} < \frac{1}{\zeta - 1} - \sum_{k<j\leq n-k} W_{j,i} = [Ww(k)]_i.$$

Therefore, $Mw(k)$ is U-shaped. Given a U-shaped vector ψ,

$$(M\psi)_n = \sum_j M_{n,j}\,\psi_j \leq \sum_j M_{n,j}\,\psi_n = \psi_n\,(Me)_n = \frac{1}{\zeta - 1}\,\psi_n.$$

Property (vi). I let $H = (-2A)^{-1}$. Then property *(vi)* is equivalent to $H_{n+1-m,1} - H_{m,1} < 1$. If $n = 2$, $H_{n+1-m,1} - H_{m,1} = 1/(2\zeta) < 1$. If $n > 2$, I define the second-order linear recurrences

$$z_k = -2\zeta z_{k-1} - z_{k-2}, \qquad k = 2, 3, \ldots, n-1$$

where $z_0 = 1$, $z_1 = 1 - 2\zeta$. I let $\zeta = \cosh\gamma$ where $\gamma > 0$. It follows from induction that

$$z_k = (-1)^k\,\frac{\cosh\left(\left(k + \frac{1}{2}\right)\gamma\right)}{\cosh\frac{\gamma}{2}} \qquad k = 0, 1, \ldots, n-1.$$

Huang and McColl (1997) calculate the entries of H in closed form. In particular,

$$H_{1,1} = \frac{1}{1 - 2\zeta - \dfrac{z_{n-2}}{z_{n-1}}} = -\frac{\cosh\left(\left(n - \frac{1}{2}\right)\gamma\right)}{2\sinh\left(n\gamma\right)\,\sinh(\gamma/2)}, \tag{7}$$

$$H_{i,1} = (-1)^{i-1}\,\frac{z_{n-i}}{z_{n-1}}\,H_{1,1} \qquad \forall i > 1.$$

It then follows that for every $1 \le i \le n$,

$$\begin{aligned} H_{n+1-i,1} - H_{i,1} &= \left(\left|\frac{z_{i-1}}{z_{n-1}}\right| - \left|\frac{z_{n-i}}{z_{n-1}}\right|\right) H_{1,1} \\ &= \frac{\cosh\left(\left(n - i + \frac{1}{2}\right)\gamma\right) - \cosh\left(\left(i - \frac{1}{2}\right)\gamma\right)}{2\sinh(n\gamma)\,\sinh\frac{\gamma}{2}} \\ &= \frac{2\sinh(n\gamma/2)\,\sinh((n-2i+1)\gamma/2)}{2\sinh(n\gamma)\,\sinh(\gamma/2)} \end{aligned} \tag{8}$$

When $i = m$, one has

$$H_{n+1-m,1} - H_{m,1} \le \frac{\sinh(n\gamma/2)\,\sinh\gamma}{\sinh\left(n\gamma\right)\,\sinh(\gamma/2)} = \frac{\cosh(\gamma/2)}{\cosh\left(n\gamma/2\right)} < 1.$$

Property (vii): Since $n\gamma$ goes to 0, it follows from (7) that $H_{1,1} \sim -1/(n\gamma^2)$. Letting $i = j = 1$ in (8), one has $H_{1,1} - H_{n,1} \sim -n/2$. This implies that

$$(\zeta - 1)\left(\max_{(i,j)} M_{ij} - \min_{(i,j)} M_{ij}\right) = (\zeta - 1)(M_{1,1} - M_{n,1}) \sim (\zeta - 1)n. \qquad \square$$

B A Microfoundation for the Search Technology

This appendix provides an example microfoundation of the search technology in Section 2. When agent i receives a demand shock at time t, the opportunity to trade is lost after an exponentially distributed time with infinitesimal[16] mean $\nu \in {}^*\mathbb{R}$. To get connected, each

[16]The hyperreals, ${}^*\mathbb{R}$, are an extension of the real numbers that contain infinite and infinitesimal numbers. An infinitesimal $\nu \in {}^*\mathbb{R}$ is a hyperreal such that $|\nu| < 1/n$, $\forall n \in \mathbb{N}$. The hyperreals are used in a branch of mathematics known as nonstandard analysis (Anderson (2000)).

line of contact of i has an independent and exponentially distributed latency time with infinitesimal mean $\eta \in {}^*\mathbb{R}$. The two infinitesimal means ν and η are "on the same order," in that neither is infinitely larger than the other. Hence, upon receiving a demand shock, the probability that i reaches one of her m quote providers before the order demand explodes is

$$\theta_m = \frac{m\eta}{m\eta + \nu},$$

which is increasing and striclty concave in $m > 0$, and $\theta_0 = 0$.

C Perfect Bayesian Equilibrium in Continuous-Time Games

I define a basic version[17] of perfect Bayesian equilibrium for continuous-time games with complete but imperfect information. Players' beliefs are given in the form of *regular conditional probabilities.* I fix a measurable space $(\Omega, \mathcal{F})$ and a filtration $(\mathcal{F}_t)_{t\geq 0}$. A *game* consists of *(i)* a finite set N of players, *(ii)* a sub-filtration $(\mathcal{F}_{it})_{t\geq 0}$ for each player $i \in N$, where the sub-σ-algebra $\mathcal{F}_{it} \subseteq \mathcal{F}_t$ represents the information available to i up to time t, *(iii)* a continuum of action spaces $(A_{it})_{t\geq 0}$ for each player i, where each A_{it} is a measurable space, *(iv)* a Markov kernel P from $\prod_{i\in N, t\geq 0} A_{it}$ to Ω, and *(v)* a utility function u_i for each player i that is a measurable function from Ω to $\mathbb{R}$. A strategy of player i is a continuum $(\sigma_{it})_{t\geq 0}$, where σ_{it} is a mapping from Ω to A_{it} that is measurable with respect to $\mathcal{F}_{it}$. A strategy profile σ is a collection of all players' strategies. The continuation value of player i at time t is $\mathrm{E}_\sigma(u_i \,|\, \mathcal{F}_{it})$, where the expectation $\mathrm{E}_\sigma(\cdot \,|\, \mathcal{F}_{it})$ is with respect to a *regular conditional probability* of $\mathrm{P} \circ \sigma$ given $\mathcal{F}_{it}$. A PBE is a strategy profile σ such that, for every $i \in N$ and $t \geq 0$, almost surely,

$$\mathrm{E}_\sigma(u_i \,|\, \mathcal{F}_{it}) \leq \mathrm{E}_{(\sigma_i', \sigma_{-i})}(u_i \,|\, \mathcal{F}_{it}).$$

[17] Other definitions typically impose additional independence restrictions on players' beliefs. Fudenberg and Tirole (1991) and Watson (2016) provide such assumptions. These restrictions are not necessary to analyze my model. Moreover, previous definitions only apply to discrete-time games.

D Proofs from Section 3

D.1 Proof of Proposition 1

It follows from expression (2) that the equilibrium spread $P^*(m)$ is strictly decreasing in the number m of dealers. When $m = 1$, the spread $P^*(1)$ satisfies the indifference condition

$$\Phi_{1,P^*(1)} = \Phi_{0,P^*(1)}.$$

When a buyside firm is isolated, its continuation value is $\Phi_{0,P} = 0$. Therefore, $\Phi_{1,P^*(1)} = 0$.

D.2 Proof of Proposition 2

I denote the dealer's indirect marginal cost of buying one unit of the asset by

$$\Delta(x) = V_{k,m,P}(x) - V_{k,m,P}(x+1)$$

Lemma 3. *Given an inventory level x, the optimal bid and offer prices of the dealer are*

$$b^*(x) = \begin{cases} -P, & \text{if } \Delta(x) \le P, \\ -P_{DP}, & \text{if } \Delta(x) > P, \end{cases} \qquad a^*(x) = \begin{cases} P, & \text{if } \Delta(x-1) \ge -P, \\ P_{DP}, & \text{if } \Delta(x-1) < -P. \end{cases}$$

Proof. When receiving a request to buy, if $V_{k,m,P}(x) - V_{k,m,P}(x-1) = -\Delta(x-1) < P$, then the trade profit P is greater than the indirect marginal cost of selling one unit of the asset. Hence, the dealer optimally posts the ask P without using its deep pocket. Conversely, if $-\Delta(x-1) > P$, the dealer is not willing to sell from its own inventory, resorting to its deep pocket to execute the trade. If $-\Delta(x-1) = P$, the dealer is indifferent, thus by the tie-breaking rule, does not use the deep pocket. The case of a request to sell is symmetric. □

Lemma 4. *I let T_1 and T_2 be two functional operators such that for every $f : \mathbb{Z} \mapsto \mathbb{R}$,*

$$T_1(f)(x) = \max\{f(x-1) + a, f(x)\} \quad \forall x \in \mathbb{Z}, \quad \text{or}$$
$$T_2(f)(x) = \max\{f(x+1) - b, f(x)\} \quad \forall x \in \mathbb{Z},$$

where $a, b \in \mathbb{R}$ are two constants. Then T_1 and T_2 preserve concavity.

Proof. If f is a concave function from $\mathbb{Z}$ to $\mathbb{R}$, then for every $x \in \mathbb{Z}$,

$$\begin{aligned} T_1(f)(x-1) - T_1(f)(x) &\leq \max\{f(x-2) - f(x-1), f(x-1) - f(x)\} \\ &\leq \min\{f(x-1) - f(x), f(x) - f(x+1)\} \\ &\leq T_1(f)(x) - T_1(f)(x+1). \end{aligned}$$

Therefore, T_1 preserves concavity. The same property holds for T_2. $\square$

Lemma 5. *The value function $V_{k,m,P}$ is even and strictly concave, in that for every $x \in \mathbb{Z}$,*

$$V_{k,m,P}(x) = V_{k,m,P}(-x), \qquad \text{and} \qquad \Delta(x) < \Delta(x+1).$$

Proof. I reparametrize the subscripts by letting $V_{\vartheta,P}$ denote $V_{k,m,P}$, where

$$\vartheta = 2\lambda \left(k\,\frac{\theta_m}{m} + \theta_{m-1} \right)$$

is the total rate of requests for quote. I write the HJB equation (3) into the following form:

$$\begin{aligned} V_{\vartheta,P}(x) &= B_{\vartheta,P,\beta}(V_{\vartheta,P})\,(x) \\ &\equiv \frac{1}{r+\vartheta} \Bigg(-\beta x^2 + \frac{\vartheta}{2} \max\{V_{\vartheta,P}(x+1) + P,\, V_{\vartheta,P}(x)\} \\ &\qquad\qquad + \frac{\vartheta}{2} \max\{V_{\vartheta,P}(x-1) + P,\, V_{\vartheta,P}(x)\} \Bigg) \end{aligned} \tag{9}$$

With a slight abuse of notation, I sometimes write B_ϑ or simply B for the Bellman operator $B_{\vartheta,P,\beta}$, and V_ϑ for $V_{\vartheta,P}$ whenever there is no ambiguity. Given two functions f, g from $\mathbb{Z}$ to $\mathbb{R}$, I write $f \leq g$ if $f(x) \leq g(x)$ for every $x \in \mathbb{Z}$. I let the space of functions

$$\Theta \equiv \left\{ f : \mathbb{Z} \to \mathbb{R} \,:\, \forall\, x, f(x) \geq -\frac{\beta}{r} x^2 \text{ and } f \leq \bar{f} \text{ for some constant } \bar{f} \in \mathbb{R} \right\}$$

be equipped with the weighted sup-norm ϱ defined as

$$\varrho(f, g) = \sup_{x \in \mathbb{Z}} \frac{|f(x) - g(x)|}{\phi(x)},$$

where $\phi(x) = \beta x^2/r + ax + b$ for some $a, b > 0$. The Bellman operator B maps the complete metric space (Θ, ϱ) into itself. For some appropriately chosen (a, b), the operator B satisfies

- (monotonicity) Given two functions $f, g \in \Theta$, if $f \leq g$, then $B(f) \leq B(g)$.
- (discounting) For every $f \in \Theta$ and $A > 0$, $B(V + A\phi) \leq B(V) + \alpha A\phi$ for some $\alpha < 1$.

The Bellman operator B is a contraction on Θ as it satisfies *Blackwell-Boyd sufficient conditions*. By the Contraction Mapping Theorem, the operator B admits a unique fixed point in Θ, which is the value function $V_{\vartheta,P}$. I let T be an operator on $\mathbb{R}^{\mathbb{Z}}$ defined by

$$T(V)(x) = \frac{\vartheta}{2}\left[\max\{V(x+1) + P,\, V(x)\} + \max\{V(x-1) + P,\, V(x)\}\right]$$

It follows from Lemma 4 that T preserves concavity. Letting $V^0(x) = -\beta x^2/r$ for every $x \in \mathbb{Z}$, then $V^0 \in \Theta$. For every $h \geq 1$, I let $V^h = B^h(V^0)$. As $h \to \infty$, one has

$$\varrho\left(V^h, V_{\vartheta,P}\right) \to 0,$$

which implies that V^h converges to $V_{\vartheta,P}$ pointwise.

It follows by induction that for every $h \geq 0$, V^h is even and strictly concave, with

$$V^h(x+1) + V^h(x-1) - 2V^h(x) \leq -\frac{2\beta}{r+\vartheta}. \tag{10}$$

Letting $h \to \infty$, one obtains that $V_{\vartheta,P}$ is even and strictly concave. $\square$

Proof of Proposition 2. I let $\bar{x}_{\vartheta,P}$ denote $\bar{x}_{k,m,P}$. I let $h \to \infty$ in (10), then for every $x \geq 0$,

$$\Delta(x) = V_{\vartheta,P}(x) - V_{\vartheta,P}(x+1) \geq \frac{(2x+1)\beta}{r+\vartheta}.$$

It follows from Lemmas 3 and 5 that the inventory threshold level $\bar{x}_{\vartheta,P}$ is such that

$$\Delta(\bar{x}_{\vartheta,P} - 1) \leq P, \qquad \Delta(\bar{x}_{\vartheta,P}) > P, \qquad \Longrightarrow \qquad \bar{x}_{\vartheta,P} < \infty. \qquad \square$$

D.3 Proofs of Lemma 1 and Proposition 3

I write $V_{\vartheta,P,\beta}$ for $V_{\vartheta,P}$ to make clear the dependence of the value function on β.

Lemma 6. *For every $x \in \mathbb{Z}$, $V_{\vartheta,P,\beta}(x)$ is jointly continuous in $(\vartheta, P, \beta) \in \mathbb{R}^{+3}$.*

Proof. First, if $0 \le \vartheta_1 \le \vartheta_2$, then $V_{\vartheta_1,P,\beta} \le V_{\vartheta_2,P,\beta}$. This is because $B^{\ell}_{\vartheta_2}(V_{\vartheta_1,P,\beta})$ converges to $V_{\vartheta_2,P,\beta}$ pointwise, and $B^{\ell+1}_{\vartheta_2}(V_{\vartheta_1,P,\beta}) \ge B^{\ell}_{\vartheta_2}(V_{\vartheta_1,P,\beta})$ for every $\ell \ge 0$ by induction. Likewise, $V_{\vartheta,P,\beta}(x)$ is non-decreasing in P and non-increasing in β for every $x \in \mathbb{Z}$.

Given a converging sequence of triples $(\vartheta_\ell, P_\ell, \beta_\ell)_{\ell \ge 0}$ of non-negative reals with some limit $(\vartheta_\infty, P_\infty, \beta_\infty)$. The sequence $(\vartheta_\ell, P_\ell, \beta_\ell)_{\ell \ge 0}$ must be bounded. For simplicity, I write V_ℓ for $V_{\vartheta_\ell,P_\ell,\beta_\ell}$ and B_ℓ for $B_{\vartheta_\ell,P_\ell,\beta_\ell}$. For every $x \in \mathbb{Z}$, the sequence $(V_\ell(x))_{\ell \ge 0}$ is bounded. Thus, there exists a subsequence $\left(V_{\varphi(\ell)}\right)_{\ell \ge 0}$ that converges pointwise to some V. For every $x \in \mathbb{Z}$,

$$\begin{aligned}
rV(x) &= \lim_{\ell\to\infty} rV_{\varphi(\ell)}(x) \\
&= \lim_{\ell\to\infty} \Bigg(-\beta_{\varphi(\ell)}x^2 + \frac{\vartheta_{\varphi(\ell)}}{2}\, \left[V_{\varphi(\ell)}(x+1) - V_{\varphi(\ell)}(x) + P_{\varphi(\ell)}\right]^+ \\
&\qquad\qquad + \frac{\vartheta_{\varphi(\ell)}}{2}\, \left[V_{\varphi(\ell)}(x-1) - V_{\varphi(\ell)}(x) + P_{\varphi(\ell)}\right]^+ \Bigg) \\
&= \qquad -\beta_\infty x^2 + \frac{\vartheta_\infty}{2}\,[V(x+1) - V(x) + P_\infty]^+ \\
&\qquad\qquad + \frac{\vartheta_\infty}{2}\,[V(x-1) - V(x) + P_\infty]^+.
\end{aligned}$$

That is, $V = B_\infty(V)$. Thus, $V = V_\infty$. Likewise, every subsequence of $(V_\ell)_{\ell \ge 0}$ admits a sub-subsequence that converges to V_∞ pointwise. The next lemma implies that V_ℓ converges to V_∞ pointwise. Thus, for every $x \in \mathbb{Z}$, $V_{\vartheta,P,\beta}(x)$ is jointly continuous in $(\vartheta, P, \beta) \in \mathbb{R}^{+3}$. $\square$

Lemma 7. *If a real sequence $(y_\ell)_{\ell \ge 0}$ is such that every subsequence of $(y_\ell)_{\ell \ge 0}$ admits a sub-subsequence that converges to the same constant $y_\infty \in \mathbb{R}$, then y_ℓ converges to y_∞.*

Proof. Otherwise, there exists some $\varepsilon > 0$ and a subsequence $\left(y_{\varphi(\ell)}\right)$ such that $\left|y_{\varphi(\ell)} - y_\infty\right| > \varepsilon$ for all ℓ. Then $\left(y_{\varphi(\ell)}\right)$ does not admit a sub-subsequence that converges to y_∞. $\square$

A function $f : \mathbb{Z} \to \mathbb{R}$ is said to be *U-shaped* if f is even and $f(x+1) > f(x)$, $\forall x \ge 0$.

Lemma 8. *The loss function $L_{k,m,P} = V_{k,m,P} - V_{k-1,m,P}$ is U-shaped.*

Proof of Lemma 8. I let $\bar{x}_\vartheta$ denote $\bar{x}_{\vartheta,P,\beta}$ and formally differentiate (3) with respect to ϑ,

$$\frac{\partial}{\partial\vartheta} V_\vartheta(x) = T_\vartheta\left(\frac{\partial}{\partial\vartheta} V_\vartheta(x)\right)(x) \tag{11}$$

$$\equiv \begin{cases} \delta\psi(x) + \frac{\delta}{2}\left[\frac{\partial}{\partial\vartheta} V_\vartheta(x+1) + \frac{\partial}{\partial\vartheta} V_\vartheta(x)\right], & x \le -\bar{x}_\vartheta, \\ \delta\psi(x) + \frac{\delta}{2}\left[\frac{\partial}{\partial\vartheta} V_\vartheta(x) + \frac{\partial}{\partial\vartheta} V_\vartheta(x-1)\right], & x \ge \bar{x}_\vartheta, \\ \delta\psi(x) + \frac{\delta}{2}\left[\frac{\partial}{\partial\vartheta} V_\vartheta(x+1) + \frac{\partial}{\partial\vartheta} V_\vartheta(x-1)\right], & |x| < \bar{x}_\vartheta, \end{cases} \tag{12}$$

where $\delta = \vartheta/(\vartheta + r)$, and for every $x \in \mathbb{Z}$,

$$\psi(x) = \frac{rV_\vartheta(x) + \beta x^2}{\vartheta^2}. \tag{13}$$

I let $\tilde{\psi} = \vartheta^2\psi/r$ be a rescaled version of ψ. It follows from (9) that for every $x \in \mathbb{Z}$,

$$\tilde{\psi}(x) = \frac{\delta}{2}\left(\left[\tilde{\psi}(x+1) - \frac{\beta(2x+1)}{r} + P\right] \vee \tilde{\psi}(x) + \left[\tilde{\psi}(x-1) + \frac{\beta(2x-1)}{r} + P\right] \vee \tilde{\psi}(x)\right).$$

It follows again from Blackwell-Boyd sufficiency conditions and the Contraction Mapping Theorem that there is a unique solution $\tilde{\psi}$ to the fixed point problem above. By induction, the function $\tilde{\psi}$ is even and convex. If $\tilde{\psi}$ is not "U-shaped," it must be that the function $\tilde{\psi}$ is constant. However, there is no constant function that solves the fixed point problem above. Therefore, $\tilde{\psi}$ is U-shaped. The function ψ is also U-shaped since ψ is a multiple of $\tilde{\psi}$.

The fixed point problem (11) admits a unique solution $\frac{\partial}{\partial\vartheta} V_\vartheta$, satisfying $A\frac{\partial}{\partial\vartheta} V_\vartheta = \psi$, where $\zeta = 1/\delta$ and A is the matrix in (6) of size $(2\bar{x}_\vartheta + 1) \times (2\bar{x}_\vartheta + 1)$. Since ψ is U-shaped,

$$\begin{aligned} &\frac{\partial}{\partial\vartheta} V_\vartheta(x) = \frac{\partial}{\partial\vartheta} V_\vartheta(-x), && \forall\, |x| \le \bar{x}_\vartheta, \\ &\frac{\partial}{\partial\vartheta} V_\vartheta(0) < \cdots < \frac{\partial}{\partial\vartheta} V_\vartheta(\bar{x}_\vartheta) \le \frac{1}{\zeta - 1}\psi(\bar{x}_\vartheta). \\ &\frac{\partial}{\partial\vartheta} V_\vartheta(x) = \frac{\psi(x) + \frac{\partial}{\partial\vartheta} V_\vartheta(x-1)/2}{(\zeta-1)+1/2}, && \forall\, x > \bar{x}_\vartheta, \end{aligned} \tag{14}$$

which follows from *(v)* of Lemma 2. One can show by induction that for every $x > \bar{x}_\vartheta$,

$$\frac{\partial}{\partial\vartheta} V_\vartheta(x-1) < \frac{1}{\zeta-1}\,\psi(x).$$

$$\Longrightarrow \frac{\partial}{\partial\vartheta} V_\vartheta(x) > \frac{(\zeta-1)\frac{\partial}{\partial\vartheta} V_\vartheta(x-1) + \frac{\partial}{\partial\vartheta} V_\vartheta(x-1)/2}{(\zeta-1)+1/2} = \frac{\partial}{\partial\vartheta} V_\vartheta(x-1).$$

Combining the inequalities above with (14), one obtains $\frac{\partial}{\partial\vartheta} V_\vartheta(x) > \frac{\partial}{\partial\vartheta} V_\vartheta(x-1)$ for every $x > 1$. Therefore, the unique solution $\frac{\partial}{\partial\vartheta} V_\vartheta$ to the fixed problem (11) is U-shaped.

I let $\vartheta_1, \vartheta_2 \geq 0$ be such that $\bar{x}_{\vartheta_1,P} = \bar{x}_{\vartheta_2,P}$, and integrate (11) over $\vartheta \in [\vartheta_1, \vartheta_2]$ to obtain

$$V_{\vartheta_2,P} - V_{\vartheta_1,P} = \int_{\vartheta_1}^{\vartheta_2} \frac{\partial}{\partial\vartheta} V_{\vartheta,P}\, d\vartheta. \tag{15}$$

Since $V_{\vartheta,P}$ is continuous with respect to ϑ (Lemma 6), then (15) holds for every $\vartheta_1, \vartheta_2 \geq 0$. Since $\frac{\partial}{\partial\vartheta} V_{\vartheta,P}$ is U-shaped for every $\vartheta \geq 0$, the function $V_{\vartheta_2,P} - V_{\vartheta_1,P}$ is also U-shaped for every $\vartheta_1, \vartheta_2 \geq 0$. In particular, the loss function $L_{k,m,P} = V_{k,m,P} - V_{k-1,m,P}$ is U-shaped. □

Since $\bar{x}_\vartheta = O\left(\vartheta^{1/3}\right)$ (Proposition 7), it then follows from property *(vii)* of Lemma 2 that

$$\frac{\partial}{\partial\vartheta} V_\vartheta(\bar{x}) - \frac{\partial}{\partial\vartheta} V_\vartheta(0) \leq \psi(\bar{x})\,\mathrm{range}\left(A^{-1}\right) = O\left((\zeta-1)\,\mathrm{range}\left(A^{-1}\right)\right) = O\left((\zeta-1)\bar{x}\right). \tag{16}$$

The inequality above will be useful in the proof of Theorem 3.

Lemma 9. *The threshold $\bar{x}_{\vartheta,P,\beta}$ is non-decreasing in $\vartheta, P \geq 0$, and non-increasing in $\beta > 0$.*

Proof. I let $\Delta_\vartheta(x) = V_\vartheta(x) - V_\vartheta(x+1)$ for every $x \in \mathbb{Z}$. Since $\frac{\partial}{\partial\vartheta} V_\vartheta$ is U-shaped, then

$$\Delta_{\vartheta_2}(x) - \Delta_{\vartheta_1}(x) = \int_{\vartheta_1}^{\vartheta_2} \frac{\partial}{\partial\vartheta} V_\vartheta(x)\, d\vartheta - \int_{\vartheta_1}^{\vartheta_2} \frac{\partial}{\partial\vartheta} V_\vartheta(x+1)\, d\vartheta < 0,$$

for every $x \in \mathbb{Z}^+$ and $\vartheta_1 < \vartheta_2$. Lemma 3 implies that $\bar{x}_{\vartheta_1,P,\beta} \leq \bar{x}_{\vartheta_2,P,\beta}$. The same technique can be applied to show that $\bar{x}_{\vartheta,P,\beta}$ is weakly decreasing in $\beta > 0$. The proof is omitted.

For each $\eta > 0$, $V_{\eta P,\eta\beta} = \eta V_{P,\beta}$ thus $\bar{x}_{\eta P,\eta\beta} = \bar{x}_{P,\beta}$. If $P_2 > P_1$, I let $\eta = P_2/P_1 > 1$, then $\bar{x}_{P_2,\beta} \geq \bar{x}_{P_2,\eta\beta} = \bar{x}_{P_1,\beta}$. When $P = 0$, $\bar{x}_{P,\beta} = 0$. Thus, $\bar{x}_{P,\beta}$ weakly increases in $P \geq 0$. □

I write $\mathcal{L}(k, m, P, \beta)$ for $\mathcal{L}(k, m, P)$ to make clear its dependence on β.

Lemma 10. *The cost $\mathcal{L}(k, m, P, \beta)$ of gouging is (a) strictly increasing in the total number k of buyside customers, (b) strictly decreasing in the total number m of dealers, (c) strictly decreasing in $\beta \in \mathbb{R}^{++}$, and (d) strictly increasing and continuous in the half spread P.*

Proof of (a) and (b). It is sufficient to establish that for every $\vartheta_2 > \vartheta_1$,

$$\frac{\partial}{\partial \vartheta} V_{\vartheta_1}(0) < \frac{\partial}{\partial \vartheta} V_{\vartheta_2}(0). \tag{17}$$

Inequality (17) would imply, for every $1 \leq k_1 < k_2$ and $1 \leq m_1 < m_2$, that

$$L_{k_1,m,P}(0) = \int_{\vartheta(k_1-1,m)}^{\vartheta(k_1,m)} \frac{\partial}{\partial \vartheta} V_\vartheta(0)\, d\vartheta < \int_{\vartheta(k_2-1,m)}^{\vartheta(k_2,m)} \frac{\partial}{\partial \vartheta} V_\vartheta(0)\, d\vartheta = L_{k_2,m,P}(0),$$

$$L_{k,m_1,P}(0) = \int_{\vartheta(k-1,m_1)}^{\vartheta(k,m_1)} \frac{\partial}{\partial \vartheta} V_\vartheta(0)\, d\vartheta < \int_{\vartheta(k-1,m_2)}^{\vartheta(k,m_2)} \frac{\partial}{\partial \vartheta} V_\vartheta(0)\, d\vartheta = L_{k,m_2,P}(0),$$

where $\vartheta(k, m) = 2k\lambda\theta_m/m$ is the reparametrization from (k, m) to ϑ.

To show (17), I formally differentiate equation (11) with respect to ϑ to obtain

$$\zeta \frac{\partial^2}{\partial \vartheta^2} V_\vartheta(x) = \begin{cases} \chi(x) + \dfrac{1}{2}\left[\dfrac{\partial^2}{\partial \vartheta^2} V_\vartheta(x+1) + \dfrac{\partial^2}{\partial \vartheta^2} V_\vartheta(x)\right], & x = -\bar{x}_\vartheta, \\[2ex] \chi(x) + \dfrac{1}{2}\left[\dfrac{\partial^2}{\partial \vartheta^2} V_\vartheta(x) + \dfrac{\partial^2}{\partial \vartheta^2} V_\vartheta(x-1)\right], & x = \bar{x}_\vartheta, \\[2ex] \chi(x) + \dfrac{1}{2}\left[\dfrac{\partial^2}{\partial \vartheta^2} V_\vartheta(x+1) + \dfrac{\partial^2}{\partial \vartheta^2} V_\vartheta(x-1)\right], & |x| < \bar{x}_\vartheta, \end{cases} \tag{18}$$

where $\zeta = 1/\delta > 1$, and

$$\chi(x) = \begin{cases} \dfrac{1}{\vartheta}\left[\dfrac{\partial}{\partial \vartheta} V_\vartheta(x+1) - \dfrac{\partial}{\partial \vartheta} V_\vartheta(x)\right], & x = -\bar{x}_\vartheta, \\[2ex] \dfrac{1}{\vartheta}\left[\dfrac{\partial}{\partial \vartheta} V_\vartheta(x-1) - \dfrac{\partial}{\partial \vartheta} V_\vartheta(x)\right], & x = \bar{x}_\vartheta, \\[2ex] \dfrac{1}{\vartheta}\left[\dfrac{\partial}{\partial \vartheta} V_\vartheta(x+1) + \dfrac{\partial}{\partial \vartheta} V_\vartheta(x-1) - 2\dfrac{\partial}{\partial \vartheta} V_\vartheta(x)\right], & |x| \leq \bar{x}_\vartheta, \end{cases}$$

Since the function $\frac{\partial}{\partial\vartheta} V_\vartheta$ is U-shaped, thus the function χ is even and

$$\sum_{\tilde{x}=-x}^{x} \chi(\tilde{x}) = \begin{cases} \frac{2}{\vartheta}\left[\frac{\partial}{\partial\vartheta} V_\vartheta(x+1) - \frac{\partial}{\partial\vartheta} V_\vartheta(x)\right] > 0, & \text{for } 0 \leq x < \bar{x}_\vartheta, \\ 0, & \text{for } x = \bar{x}_\vartheta. \end{cases} \tag{19}$$

The linear system (18) can be written as $A \frac{\partial^2}{\partial\vartheta^2} V_\vartheta = \chi$. Then $\frac{\partial^2}{\partial\vartheta^2} V_\vartheta = A^{-1}\chi$. In particular,

$$\frac{\partial^2}{\partial\vartheta^2} V_\vartheta(0) = \sum_{x=-\bar{x}_\vartheta}^{\bar{x}_\vartheta} A^{-1}_{0,x} \cdot \chi(x) = \sum_{x=0}^{\bar{x}_\vartheta - 1} \left(A^{-1}_{0,x} - A^{-1}_{0,x+1}\right) \sum_{\tilde{x}=-x}^{x} \chi(\tilde{x}) > 0.$$

The last inequality follows from (19) and property *(ii)* of Lemma 2.

Given $\vartheta_1, \vartheta_2 \geq 0$ such that $\bar{x}_{\vartheta_1} = \bar{x}_{\vartheta_2}$, I integrate equation (18) over $\vartheta \in [\vartheta_1, \vartheta_2]$ to obtain

$$\frac{\partial}{\partial\vartheta} V_{\vartheta_2} - \frac{\partial}{\partial\vartheta} V_{\vartheta_1} = \int_{\vartheta_1}^{\vartheta_2} \frac{\partial^2}{\partial\vartheta^2} V_\vartheta \, d\vartheta.$$

Since $\bar{x}_\vartheta$ is non-decreasing in ϑ (Lemma 9), the set of discontinuity points of $\bar{x}_\vartheta$ is discrete and admits no accumulation point. I let ϑ_0 be a discontinuity point of $\bar{x}_\vartheta$, and define $\bar{x}^-$ as

$$\bar{x}^- \equiv \lim_{\vartheta \uparrow \vartheta_0} \bar{x}_\vartheta \leq \bar{x}_{\vartheta_0}.$$

The same argument used in the proof of Lemma 6 implies that for every $x \in \mathbb{Z}$,

$$\lim_{\vartheta \uparrow \vartheta_0} \frac{\partial}{\partial\vartheta} V_\vartheta(x) = \widetilde{V}(x),$$

where the function $\widetilde{V}$ is the unique solution to the fixed point problem

$$\widetilde{V}(x) = \begin{cases} \delta\psi(x) + \frac{\delta}{2}\left[\widetilde{V}(x+1) + \widetilde{V}(x)\right], & x \leq -\bar{x}^-, \\ \delta\psi(x) + \frac{\delta}{2}\left[\widetilde{V}(x) + \widetilde{V}(x-1)\right], & x \geq \bar{x}^-, \\ \delta\psi(x) + \frac{\delta}{2}\left[\widetilde{V}(x+1) + \widetilde{V}(x-1)\right], & |x| < \bar{x}^-, \end{cases}$$

I let $\widetilde{V}^0 = \widetilde{V}$, and $\widetilde{V}^{h+1} = T_{\vartheta_0}(\widetilde{V}^h)$, where T_{ϑ_0} is defined in (12). Then $\widetilde{V}^h$ converges to

$\frac{\partial}{\partial\vartheta} V_{\vartheta_0}$ pointwise as $h \to \infty$. Since $\bar{x}^- \leq \bar{x}_{\vartheta_0}$ and $\widetilde{V}^0$ is U-shaped, one has

$$\widetilde{V}^1(x) \begin{cases} = \widetilde{V}^0(x) & |x| < \bar{x}^-, \\ > \widetilde{V}^0(x) & \bar{x}^- \leq |x| < \bar{x}_{\vartheta_0}, \\ = \widetilde{V}^0(x) & |x| \geq \bar{x}_{\vartheta_0}. \end{cases}$$

Thus, $\widetilde{V}^0 \leq \widetilde{V}^1$. It then follows that $\widetilde{V}^h \leq \widetilde{V}^{h+1}$ for every $h \geq 0$. Letting $h \to \infty$, one has

$$\lim_{\vartheta \uparrow \vartheta_0} \frac{\partial}{\partial\vartheta} V_\vartheta = \widetilde{V} \leq \frac{\partial}{\partial\vartheta} V_{\vartheta_0}.$$

Hence,
$$\frac{\partial}{\partial\vartheta} V_{\vartheta_2}(0) - \frac{\partial}{\partial\vartheta} V_{\vartheta_1}(0) \geq \int_{\vartheta_1}^{\vartheta_2} \frac{\partial^2}{\partial\vartheta^2} V_\vartheta(0)\, d\vartheta > 0.$$

for every $\vartheta_2 > \vartheta_1$. This establishes (17) and completes the proof of *(a)* and *(b)* of Lemma 10.

Part (c). The same technique used in the proof of parts *(a)* and *(b)* can be applied to show part *(c)*. One should apply the continuity property of $V_{k,d,P,\beta}$ in β (Lemma 6) and the monotonicity of $\bar{x}_{k,d,P,\beta}$ in β (Lemma 9). The proof is omitted.

Part (d). For every $\eta > 0$, $V_{k,m,\eta P,\eta\beta} = \eta V_{k,m,P,\beta}$. If $P_2 > P_1$, I let $\eta = P_2/P_1 > 1$, then

$$L_{k,m,P_2,\beta}(0) > L_{k,m,P_2,\eta\beta}(0) = \eta\, L_{k,m,P_1,\beta}(0) > L_{k,m,P_1,\beta}(0).$$

That is, $L_{k,m,P,\beta}(0)$ is strictly increasing in $P \in \mathbb{R}^+$. The continuity of $L_{k,m,P,\beta}(0)$ in P is implied by that of $V_{k,m,P}(0)$ as per Lemma 6. □

Proof of Lemma 1. Since $\mathcal{L}(k, m, P)$ is continuous and weakly increasing in P (Lemma 10), condition (4) is equivalent to $P \geq \underline{P}(k, m)$. □

Proposition 3 follows immediately from *(a)* and *(b)* of Lemma 10.

D.4 Proof of Theorem 1

Step 1: I first show that $\sigma^*(m)$ is a PBE for every $m \leq m^*$. When receiving a demand shock, the expected gain from search is $\theta_m[\pi - P^*(m)]$ for a buyside firm and $\theta_{m-1}[\pi - P^*(m)]$ for a dealer. Without an exogenous need to trade, the expected gain from search is 0. Hence, the search strategy S^* is optimal for every agent. Since the total gain per trade is π, the response strategy ρ^* is optimal. If an agent has d dealer accounts, the rate of net benefit for the agent is $r\Phi_{d,P^*(m)}$. It is thus strictly optimal to have $m-1$ or m dealer accounts. Hence, the account maintenance strategy N_m^* is optimal. If a buyside firm i receives a request for quote at time t, given that the associated quote seeker's equilibrium strategy is to discontinue its trading account with i immediately after time t, the quoting strategy p_0^* is thus optimal for i. Finally, since $m \leq m^*$ implies the condition $\Pi(P^*(m)) \leq \mathcal{L}(n-m, m, P^*(m))$ for no gouging, each dealer has no incentive to gouge by the One-Shot Deviation Principle. The quoting strategy $p^*(m)$, determined by dealer's HJB equation (3), is thus optimal for dealers.

Step 2: I suppose that $G(m)$ is an equilibrium network for some integer $m \geq 0$, and that $\sigma = (S, p, \rho, N^{\text{out}})$ is a supporting equilibrium of $G(m)$. I show that on the equilibrium path, the ask price of any given dealer $j \in J$ is some constant a_j^*, and its bid price is some constant b_j^*. Formally, for every busyide firm i, every dealer j and every integer $\ell \geq 1$, I let $\tau_{i\ell}$ denote the time of the ℓ'th request for quote of i, and $\tau_{j\ell}$ denote the time of dealer j providing the ℓ'th quote. I use the subscripts $i\ell$, $j\ell$ and $i\ell^-$ to denote "at time $\tau_{i\ell}$," "at time $\tau_{j\ell}$," and "right before time $\tau_{i\ell}$" respectively. I show that for every $\ell \geq 1$, $a_{j\ell} = a_j^*, b_{j\ell} = b_j^*$ almost surely.

Since the search strategy S_i of every given agent i is stationary, and i searches only a finite number of times during any finite time interval on the equilibrium path, it must be that i searches only upon receiving a demand shock. That is, $S_i = S^*$ for every agent i. Conditional on successfully reaching a quote provider, it must be that agent i accepts any ask $a < \pi$ and any bid $b > -\pi$. For every $j \in J$, $a \in \mathbb{R}$ and $\ell = 1, 2, \ldots$, I let

$$A_{i\ell}(a, j) = \{O_{i\ell} = \text{Buy}, \tilde{a}_{i\ell} = a, j_{i\ell} = j\}$$

denote the event that buyside firm i receives an ask price a from dealer j at time $\tau_{i\ell}$, and

$$a_j^* = \sup\{a \in [-\pi, \pi] : \ \forall i \in I, N_{i\ell}^{\text{out}} = J \text{ almost surely on the event } A_{i\ell}(a, j)\} \tag{20}$$

be the highest ask price that j may offer without triggering account termination by any buyside firm on the equilibrium path. Likewise, for every $b \in \mathbb{R}$, I let

$$B_{i\ell}(b, j) = \left\{ O_{i\ell} = \text{Sell}, \tilde{b}_{i\ell} = b, j_{i\ell} = j \right\},$$

and $\quad b_j^* = \sup\{b \in [-\pi, \pi] \,:\, \forall i \in I, N_{i\ell}^{\text{out}} = J \text{ almost surely on the event } B_{i\ell}(a, j)\}$

be the lowest bid price that j may post without triggering account termination.

If dealer j posts some ask price $a_{j\ell} < a_j^*$, then there exists some $\varepsilon > 0$ such that $a + \varepsilon \leq \pi$, and offering an ask price of $a_{j\ell} + \varepsilon$ would not trigger any account termination. Dealer j is thus strictly better off if it raises its ask price by ε, contradicting the optimality of its quoting strategy. If $a_{j\ell} > a_j^*$, then by definition (20) of a_j^*, the ask price $a_{j\ell}$ triggers at least one account termination by some buyside firm with some positive probability. This contradicts that $G(m)$ is the equilibrium trading network. Therefore, the ask price of dealer j is always a_j^* on the equilibrium path. Likewise, the bid price of dealer j is always b_j^*. Consequently, the ask a_j^* and the bid b_j^* do not trigger account termination by any agent.

Step 3: I next show that for every dealer $j \in J$, the bid-ask spread is $a_j^* - b_j^* = 2P^*(m)$. The equilibrium continuation utility of any buyside firm i at any given time $t \geq 0$ is

$$\Phi = \frac{\lambda}{r} \frac{\theta_m}{m} \sum_{j \in J} \left(\pi - a_j^* + \pi + b_j^* \right) - \frac{mc}{r}.$$

If i terminates its trading account with a given dealer $j \in J$, its continuation utility is

$$\Phi_{-j} = \frac{\lambda}{r} \frac{\theta_{m-1}}{m-1} \sum_{j' \in J/\{j\}} \left(\pi - a_{j'}^* + \pi + b_{j'}^* \right) - \frac{(m-1)c}{r}.$$

It must be that i is indifferent to maintaining all its dealer accounts or terminating one of them. That is, for every dealer $j \in J$, $\Phi = \Phi_{-j}$. It then follows that the spread $a_j^* - b_j^* = 2P_j^*$ is the same for every dealer $j \in J$. To see this, I let $j_1 = \text{argmax}_j\, P_j^*$ and $j_2 = \text{argmax}_j\, P_j^*$ be the dealers posting the largest and the smallest spread respectively. If $P_{j_1}^* > P_{j_2}^*$, then terminating the trading account with j_1 is strictly better than terminating the account with

j_2. That is, $\Phi_{-j_1} > \Phi_{-j_2}$, which contradicts $\Phi = \Phi_{-j_1} = \Phi_{-j_2}$. Thus, there exists some constant P^* such that $P^*_j = P^*$ for every dealer $j \in J$. The equilibrium continuation utility of a buyside firm is then $\Phi = \Phi_{m,P^*}$, where $\Phi_{d,P}$ is given by (2). The continuation utility of a buyside firm after terminating one dealer account is Φ_{m-1,P^*}. The indifference condition $\Phi_{m,P^*} = \Phi_{m-1,P^*}$ implies $P^* = P^*(m)$, where $P^*(m)$ is the equilibrium spread given by (2).

Step 4: I suppose $G(m)$ is an equilibrium network for some given $m \geq 0$. For a given dealer, I let $a^* = P^*(m) + h$ and $b^* = -P^*(m) + h$ be its equilibrium ask and bid prices respectively, for some $h \geq 0$. The dealer's value function $V_{k,h}$ solves the following HJB equation:

$$\begin{aligned} rV_{k,h}(x) = -\beta x^2 &+ \lambda \left(k\,\frac{\theta_m}{m} + \theta_{m-1} \right) [V_{k,h}(x+1) - V_{k,h}(x) + P^*(m) - h]^+ \\ &+ \lambda \left(k\,\frac{\theta_m}{m} + \theta_{m-1} \right) [V_{k,h}(x-1) - V_{k,h}(x) + P^*(m) + h]^+. \end{aligned} \tag{21}$$

This HJB equation differs from (3) only in the bid-ask quotes. It follows from the Blackwell's sufficient conditions and the Contraction Mapping Theorem that there is a unique solution $V_{k,h}$ to (21) that is bounded above by some constant and below by $-\beta x^2/r$. I let $y_h = V_{k,h} + \beta x^2/r$. It then follows from the HJB equation (21) that for every $x \in \mathbb{Z}$,

$$\begin{aligned} y_h(x) = \frac{\delta}{2} \Bigg(&\left[y_h(x+1) - \frac{\beta(2x+1)}{r} + P - h \right] \vee y_h(x) \\ &+ \left[y_h(x-1) + \frac{\beta(2x-1)}{r} + P + h \right] \vee y_h(x) \Bigg), \end{aligned} \tag{22}$$

I extend the domain of the function y_h from $\mathbb{Z}$ to $\mathbb{R}$. That is, (22) holds for every $x \in \mathbb{R}$. The Contraction Mapping Theorem implies that the function y_h is uniquely determined by (22). It follows from value iteration that the function y_0 is even and convex. Hence, it must be that y_0 is a continuous function on $\mathbb{R}$. It can be verified that for every $x \in \mathbb{R}$,

$$y_h(x) = y_0 \left(x + \frac{rh}{2\beta} \right).$$

That is, y_h is obtained by simply shifting the function y_0 to the left by $rh/(2\beta)$. I extend the domain of the function $V_{k,h}$ from $\mathbb{Z}$ to $\mathbb{R}$ by defining $V_{k,h} = y_h - \beta x^2/r$. Hence, $y_h(x)$ and thus $V_{k,h}(x)$ are jointly continuous in (x, h). The Intermediate Value Theorem implies the existence of some $\tilde{x}_{k,0} \in \mathbb{R}$ such that

$$\bar{x}_k = \lfloor \tilde{x}_{k,0} \rfloor, \qquad V_{k,0}(\tilde{x}_{k,0} - 1) - V_{k,0}(\tilde{x}_{k,0}) = P^*(m).$$

I next consider the dealer's incentive to gouge. The one-shot benefit of gouging is

$$\Pi(h) = \max\{\pi - a^*, \pi + b^*\} = \pi - P^*(m) + h.$$

If the dealer gouges, the probability that it loses a buyside customer is *at most* $k/(k+m-1)$. (this probability could be lower since some buyside firms may choose not to terminate their accounts). I let $L_{k,h} = V_{k,h} - V_{k-1,h}$. Since the dealer optimally controls its inventory within the interval $I_h = \left[-\tilde{x}_{k,0} - \frac{rh}{2\beta}, \tilde{x}_{k,0} - \frac{rh}{2\beta}\right]$, a necessary condition for no gouging is given by

$$\Pi(h) \leq \mathcal{L}(k, h) \equiv \frac{k}{k+m-1} \min_{x \in I_h} L_{k,h}(x). \tag{23}$$

I will show that, for every $h \geq 0$,

$$\mathcal{L}(k, h) \leq \mathcal{L}(k, 0) + h. \tag{24}$$

Then condition (23) would imply $\Pi(0) \leq \mathcal{L}(k, 0)$, which is equivalent to $m \leq m^*$ when $k = n - m$. This would complete *Step 4*. Since $\mathcal{L}(k, h)$ is even and periodic in h with period $2\beta/r$. Hence, it suffices to show that (24) holds for every $h \leq \beta/r$. For every $h \leq \beta/r$,

$$\min_{x \in I_h} L_{k,h}(x) = L_{k,h}(0).$$

It is therefore sufficient to show, for every $h \leq \beta/r$, that

$$L_{k,h}(0) \leq L_{k,0}(0) + h. \tag{25}$$

I formally differentiate $V_{k,h}(x)$ with respect to h in (21), to obtain

$$\zeta \frac{\partial}{\partial h} V_{k,h}(x) = \begin{cases} \frac{1}{2}\left[\frac{\partial}{\partial h} V_{k,h}(x+1) + \frac{\partial}{\partial h} V_{k,h}(x-1)\right], & \tilde{x}_{k,h} + 1 \leq x + \frac{rh}{2\beta} \leq \tilde{x}_{k,h} - 1, \\ \frac{1}{2}\left[\frac{\partial}{\partial h} V_{k,h}(x+1) + \frac{\partial}{\partial h} V_{k,h}(x) - 1\right], & x + \frac{rh}{2\beta} < \tilde{x}_{k,h} + 1, \\ \frac{1}{2}\left[\frac{\partial}{\partial h} V_{k,h}(x-1) + \frac{\partial}{\partial h} V_{k,h}(x) + 1\right], & x + \frac{rh}{2\beta} > \bar{x}_{k,h} - 1. \end{cases}$$

I let ℓ be the number of integers in the interval I_h, $s = \lfloor (\ell+1)/2 \rfloor$, and ϖ be the vector

$$\varpi = (-1/2, 0, \dots, 0, 1/2)^\top.$$

The linear system can be written as $A \frac{\partial}{\partial h} V_{k,h} = \varpi$, where A is the matrix (6) of size $\ell \times \ell$.

For every $h \leq \beta/r$, it follows from properties *(iii)*, *(v)* and *(vii)* of Lemma 2 that

$$0 \leq \frac{\partial}{\partial h} V_{k,h}(0) = \frac{1}{2}\left(-A^{-1}_{\ell+1-s,1} + A^{-1}_{\ell+1-s,\ell}\right) = \frac{1}{2}\left(-A^{-1}_{\ell+1-s,1} + A^{-1}_{s,1}\right) \leq 1.$$

$$\Longrightarrow L_{k,h}(0) = L_{k,h}(0) + [V_{k,h}(0) - V_{k,0}(0)] - [V_{k,h}(0) - V_{k-1,0}(0)] \leq L_{k,h}(0) + h.$$

D.5 Proof of Theorem 2 and Corollary 1

The proof of Theorem 2 closely parallels that of Theorem 1, and is thus omitted. In the equilibrium network G, the sum of outdegrees must equal to the sum of indegrees. Thus,

$$m^* n - \frac{\sum_{j \in J} \ell_j}{m-1} = \sum_{j \in J} (k_j + \ell_j).$$

Since $k_j \geq k(m^*, \ell_j)$ for every $j \in J$ (Theorem 2), it follows from the next lemma that

$$m^* n \geq \sum_{j \in J} \left(k(m^*, \ell_j) + \frac{m}{m-1} \ell_j \right) > |J| \, [k(m^*, 0) - 1],$$

which implies

$$|J| < \frac{m^* n}{k(m^*, 0) - 1}.$$

Lemma 11. *For every integers $m > 0$ and $\ell \geq 0$,*

$$k(m,\ell) + \frac{m}{m-1}\,\ell > k(m,0) - 1.$$

Proof. It follows from the definition of $k(m,\ell)$ that

$$\Pi(P^*(m)) > \mathcal{L}(k(m,0) - 1, m, 0, P^*(m)). \tag{26}$$

Let $\vartheta = 2\lambda(k\theta_m/m + \ell\theta_{m-1}/(m-1))$. With reparametrization, I write $V_{\vartheta,P}$ for $V_{k,m,\ell,P}$, and $L_{\vartheta,P}$ for $L_{k,m,\ell,P}$. It follows from (15) and (17) in the proof of Lemma 10 that $L_{\vartheta,P}(0)$ is strictly increasing in $\vartheta \geq 0$. It then follows that

$$\mathcal{L}(k(m,0) - 1, m, 0, P^*(m)) \geq \mathcal{L}\left(\left\lfloor k(m,0) - 1 - \frac{m}{m-1}\,\ell \right\rfloor, m, \ell, P^*(m)\right).$$

Combining with (26), one has

$$\Pi(P^*(m)) > \mathcal{L}\left(\left\lfloor k(m,0) - 1 - \frac{m}{m-1}\,\ell \right\rfloor, m, \ell, P^*(m)\right).$$

Therefore, $$k(m,\ell) \geq \left\lfloor k(m,0) - \frac{m}{m-1}\,\ell \right\rfloor > k(m,0) - 1 - \frac{m}{m-1}\,\ell.$$ □

D.6 Proof of Proposition 4

Step 1: If σ is a supporting equilibrium for some network G with $|J(G)| > \mu(G)$, I let $m = \mu(G)$. Since the equilibrium spread is $2P^*(m)$ (Theorem 1), the quotes of every given dealer j are $a_j^* = P^*(m) + h_j$ and $b_j^* = P^*(m) + h_j$ for some h_j. I let E be the set of dealers with the $J(G) - m$ smallest mid-quotes $|h_j|$ in magnitude. These dealers are the ones who "exit," in that they give up all trading accounts they host to the remaining m dealers in $J' = J(G)\backslash E$ and become buyside firms. Specifically, if a given agent has m dealer accounts, then I replace each of his accounts with dealers in E by an account with one of the m dealers in J'. Among the m dealers in J', I let j' be the dealer with the smallest mid-quote $|h_j|$. If a given agent has $m-1$ dealer accounts, then I replace each of his accounts with dealers in E by an account with one of the $m-1$ dealers in $J'\backslash\{j'\}$. I let $\vartheta_j = 2\lambda[k_j\theta_m/m + \ell_j\theta_{m-1}/(m-1)]$

and $\vartheta'_j = 2\lambda \left[k'_j\theta_m/m + \ell'_j\theta_{m-1}/(m-1)\right]$. I define a strategy profile σ' that differs from σ only in the quoting strategies of the m dealers, in that each dealer j posts the same bid-ask quotes a^*_j and b^*_j as in σ, but with the inventory thresholds determined by the rate ϑ'_j instead of ϑ_j. Then the value of j from serving requests for quote is given by $V'_j = V_{\vartheta'_j,h_j}(0)$. With the same proof of Lemma 10, one can show that $V_{\vartheta,h}(0)$ is increasing and strictly convex in ϑ for every $h \in \mathbb{R}$. Since $\sum_{j\in J(G)} \vartheta_j = \sum_{j\in J(G)} \vartheta'_j$, then

$$\sum_{j\in J(G)} V'_j > \sum_{j\in J(G)} V_j.$$

by Jensen's inequality, where $V_j = V_{\vartheta_j,h_j}(0)$. Then $U(\sigma') > U(\sigma)$.

Step 2: From G', I let all buyside firms without an account with j' open one with j', if

$$\frac{\theta_m}{m}\frac{\partial}{\partial\vartheta} V'_{j'} \geq \left(\frac{\theta_{m-1}}{m-1} - \frac{\theta_m}{m}\right)\frac{\partial}{\partial\vartheta}\sum_{j\neq j'} V'_j. \tag{27}$$

The resulting network is the concentrated core-periphery network $G(m)$. I let σ'' differ from σ' only in the quoting strategy of j', in that j' posts the bid-ask quotes $a^*_{j'}$ and $b^*_{j'}$ with the inventory thresholds determined by the rate $\vartheta_{n-m,m}$ instead of ϑ'. Then σ'' is a supporting equilibrium of $G(m)$, and $U_{\sigma''} \geq U_{\sigma'} > U_\sigma$. If inequality (27) does not hold, then one can construct a supporting equilibrium σ'' for $G(m-1)$ in a similar way, such that $U_{\sigma''} > U_\sigma$.

E Proofs from Section 4

E.1 Proof of Proposition 6

Part (i): I fix $m \geq 1$ and $P > 0$, and suppress m and P from the subscripts to simplify notations. For example, V_n denotes $V_{n-m,m,P}$. Since $\frac{\partial}{\partial\vartheta} V_\vartheta$ is U-shaped, then for every $x \geq 0$,

$$\begin{aligned}&[V_{n+1}(x) - V_{n+1}(x+1)] - [V_n(x) - V_n(x+1)]\\ &= \int_{\vartheta(n-m,m)}^{\vartheta(n-m+1,m)} \left[\frac{\partial}{\partial\vartheta} V_\vartheta(x) - \frac{\partial}{\partial\vartheta} V_\vartheta(x+1)\right] d\vartheta < 0.\end{aligned}$$

Since the sequence $[V_n(x) - V_n(x+1)]_{n\geq m}$ is strictly decreasing, it admits some limit $\Delta_\infty(x)$.

It will be shown in the proof of Proposition 7 that $\bar{x}_n \to \infty$ as $n \to \infty$. Then for every $x \in \mathbb{Z}^+$, $x < \bar{x}_n$ for n sufficiently large, and it follows from (9) that

$$\begin{aligned} & rV_n(x) \\ = \quad & -\beta x^2 + \lambda\left((n-m)\frac{\theta_m}{m} + \theta_{m-1}\right)[V_n(x+1) + V_n(x-1) - 2V_n(x) + 2P] \\ \sim \quad & n\lambda\frac{\theta_m}{m}[\Delta_\infty(x-1) - \Delta_\infty(x) + 2P] \end{aligned} \tag{28}$$

Where the symbol $\sim$ indicates asymptotic equivalence as $n \to \infty$. Letting $x = 0$, one has

$$rV_n(0) \sim n\lambda\frac{\theta_m}{m}[-2\Delta_\infty(0) + 2P]. \tag{29}$$

For every $x \geq 0$,

$$r[V_n(0) - xP] \leq rV_n(x) \leq rV_n(0) \quad \Longrightarrow \quad rV_n(x) \sim n\lambda\frac{\theta_m}{m}[-2\Delta_\infty(0) + 2P]. \tag{30}$$

By comparing the asymptotic equivalences in (28) and (30), one obtains

$$\Delta_\infty(x) - \Delta_\infty(x-1) = 2\Delta_\infty(0),$$

for every $x \in \mathbb{Z}^+$. Thus

$$\Delta_\infty(x) = (2x+1)\Delta_\infty(0).$$

If $\Delta_\infty(0) > 0$, then $\Delta_\infty(x) > P$ for $x > [P/\Delta_\infty(0) - 1]/2$, which implies $\bar{x}_n < [P/\Delta_\infty(0) - 1]/2$. The last inequality contradicts with the fact that $\bar{x}_n$ goes to infinity as $n \to \infty$. Therefore, $\Delta_\infty(0) = 0$. It then follows from (29) that

$$rV_n(0) \sim 2n\lambda\frac{\theta_m}{m}P. \tag{31}$$

Since $L_n(0)$ is strictly increasing in $n \geq m$ (Lemma 10), it has a (possibly infinite) limit as $n \to \infty$. It then follows Cesàro's Theorem that

$$\frac{V_n(0)}{n-m} = \frac{\sum_{k=m+1}^{n} L_k(0)}{n-m} \xrightarrow{n\to\infty} \lim_{n\to\infty} L_n(0).$$

The equivalence in (31) then implies that

$$\lim_{n\to\infty} L_n(0) = \frac{2\lambda\theta_m}{rm}\, P$$

One can then solve $\Pi(P) = \mathcal{L}(n-m, m, P)$ to obtain

$$\lim_{n\to\infty} \underline{P}(n-m, m) = \frac{mr\pi}{2\lambda\theta_m + mr}.$$

Since $\mathcal{L}(n-m, m, P)$ is strictly increasing in n, the sustainable spread $\underline{P}(n-m, m)$ is strictly decreasing in n. Thus, the limiting number m^*_∞ of dealers is the largest integer m such that

$$\frac{mr\pi}{2\lambda\theta_m + mr} = \lim_{n\to\infty} \underline{P}(n-m, m) < P^*(m).$$

Part (ii) (dependence of m^* on π): To indicate the dependence of endogenous variables on the parameter π, I will write $P^*(m,\pi)$ for the equilibrium spread, and $\underline{P}(m,\pi)$ for the sustainable dealer spread. The loss function $\mathcal{L}(k, m, P)$ does not depend on π.

Given some $\pi_1 < \pi_2$, one has, for every $m \geq 1$, $P > 0$, and $\ell = 1, 2$,

$$\pi_\ell - \underline{P}(m, \pi_\ell) = \mathcal{L}(n-m, m, \underline{P}(m, \pi_\ell)).$$

Since the loss $\mathcal{L}(n-m, m, P)$ is strictly increasing in P (Lemma 10), it must be that

$$\pi_1 - \underline{P}(m, \pi_1) < \pi_2 - \underline{P}(m, \pi_2), \qquad \underline{P}(m, \pi_2) > \underline{P}(m, \pi_1).$$

That is, when the total gain per trade increases from π_1 to π_2, the increase in the dealer sustainable spread $\underline{P}(m,\pi)$ is strictly less than $\pi_2 - \pi_1$. On the other hand, one has

$$P^*(m, \pi_2) - P^*(m, \pi_1) = \pi_2 - \pi_1.$$

That is, the equilibrium spread increases more than the sustainable spread. Therefore, the core size m^* is weakly increasing in π.

Part (ii) (dependence of m^* on λ): The same technique used in the proof of Lemma 10 can be applied to show that the loss $\mathcal{L}(k,m,P)$ from gouging is strictly increasing in λ, for every $k \geq 1$ and $P > 0$. Hence, the dealer-sustainable spread $\underline{P}(m)$ is strictly decreasing in λ. On the other hand, the equilibrium spread $P^*(m)$ is strictly increasing in λ (see expression (2)). Therefore, the core size m^* is weakly increasing in λ.

Part (ii) (dependence of m^* on c): As c decreases, $P^*(m)$ increases, while $\underline{P}(m)$ is not affected. The core size m^* thus weakly increases.

Part (ii) (dependence of m^* on β): For every $m \geq 1$, $k \geq 1$ and $P > 0$, the loss function $\mathcal{L}(k,m,P,\beta)$ is strictly decreasing in $\beta \in \mathbb{R}^{++}$ (Lemma 10). Thus, the dealer sustainable spread $\underline{P}(m)$ is strictly increasing in β. However, the equilibrium spread $P^*(m)$ does not depend on β. Therefore, the core size m^* is weakly decreasing in β.

Part (iii): When β increases or n decreases, the equilibrium number m^* of dealers weakly decreases. With less competition, dealers widen their equilibrium spread offer. This can be seen directly from expression (2) of the equilibrium spread $P^*(m^*)$.

E.2 Proof of Proposition 7

Lemma 9 shows that $\bar{x}_{n-m,m,P^*(m)}$ is weakly decreasing in β and weakly increasing in $n\lambda$. To drive the desired asymptotic, I fix some $m \geq 1$ and $P > 0$, and let $\vartheta = 2\lambda((n-m)\theta_m/m + \theta_{m-1})$. With reparametrization, I write V_ϑ for $V_{n-m,m,P}$ and $\bar{x}_\vartheta$ for $\bar{x}_{n-m,m,P}$. It is sufficient to show that $\bar{x}_\vartheta = \Theta\left(\vartheta^{1/3}\right)$ as ϑ goes to infinity. It follows from (3) that

$$V_\vartheta(x) = T_1(V_\vartheta)(x), \qquad -\bar{x}_\vartheta < x < \bar{x}_\vartheta. \tag{32}$$

$$V_\vartheta(x) = T_2(V_\vartheta)(x), \qquad x \geq \bar{x}_\vartheta. \tag{33}$$

where for every function $V : \mathbb{Z} \to \mathbb{R}$,

$$T_1(V)(x) = \frac{1}{\vartheta + r}\left(-\beta x^2 + \frac{\vartheta}{2}\left[V(x-1) + V(x+1) + 2P\right]\right)$$

$$T_2(V)(x) = \frac{1}{\vartheta + r}\left(-\beta x^2 + \frac{\vartheta}{2}\left[V(x-1) + V(x) + P\right]\right).$$

A quadratic solution U^0_ϑ of (32) is given by

$$U^0_\vartheta(x) = -\frac{\beta}{r}x^2 + \frac{\vartheta}{r}\left(P - \frac{\beta}{r}\right).$$

To obtain all solutions of (32), I consider its homogeneous version:

$$rV(x) = \frac{\vartheta}{2}\left[V(x-1) + V(x+1) - 2V(x)\right]. \tag{34}$$

The set of solutions to the difference equation above forms a 2-dimensional vector space

$$\{ae^{d_\vartheta x} + \tilde{a}e^{d_\vartheta x} \,:\, a, \tilde{a} \in \mathbb{R}\}, \qquad \text{where} d_\vartheta = \sqrt{\frac{2r}{\vartheta}} + O\left(\vartheta^{-\frac{3}{2}}\right).$$

Therefore, the solutions to (32) are

$$\mathbb{Z} \ni x \mapsto -\frac{\beta}{r}x^2 + \frac{\vartheta}{r}\left(P - \frac{\beta}{r}\right) + ae^{d_\vartheta x} + \tilde{a}e^{-d_\vartheta x},$$

where $a, \tilde{a} \in \mathbb{R}$. The value function V_ϑ must be equal to one of the solutions U_ϑ in the region $-\bar{x}_\vartheta \le x \le \bar{x}_\vartheta$, for some $a = a_\vartheta$ and $\tilde{a} = \tilde{a}_\vartheta$. Since the function V_ϑ is even, one must have $a_\vartheta = \tilde{a}_\vartheta$. Hence, for every integer $x \in [-\bar{x}_\vartheta, \bar{x}_\vartheta]$,

$$V_\vartheta(x) = U_\vartheta(x) \equiv -\frac{\beta}{r}x^2 + \frac{\vartheta}{r}\left(P - \frac{\beta}{r}\right) + a_\vartheta \cosh(d_\vartheta x). \tag{35}$$

Solving equation (33), one obtains, for every integer $x \ge \bar{x}_\vartheta - 1$,

$$\begin{aligned} V_\vartheta(x) = W_\vartheta(x) &\equiv W^0_\vartheta(x) + b_\vartheta e^{c_\vartheta x} \\ &\equiv -\frac{\beta}{r}x^2 + \frac{\vartheta}{r}\frac{\beta}{r}x - \left(\frac{\vartheta}{r}\right)^2\frac{\beta}{2r} + \frac{\vartheta}{2r}\left(P - \frac{\beta}{r}\right) + b_\vartheta e^{c_\vartheta x}, \end{aligned} \tag{36}$$

for some $b_\vartheta \in \mathbb{R}$, where $\qquad c_\vartheta = -\frac{2r}{\vartheta} + 2\left(\frac{r}{\vartheta}\right)^2 + O\left(\vartheta^{-3}\right).$

I show that the undetermined coefficients a_ϑ and b_ϑ are non-negative. For this purpose, I define V^0_ϑ as an even function from $\mathbb{Z}$ to $\mathbb{R}$ such that for every $x \in \mathbb{Z}^+$,

$$V^0_\vartheta(x) = \max\left\{U^0_\vartheta(x), W^0_\vartheta(x)\right\}.$$

I let B_ϑ be the Bellman operator defined in (9). Then one has

$$\begin{cases} B_\vartheta\left(V_\vartheta^0\right) \geq T_1\left(V_\vartheta^0\right) \geq T_1\left(U_\vartheta^0\right) = U_\vartheta^0, \\ B_\vartheta\left(V_\vartheta^0\right) \geq T_2\left(V_\vartheta^0\right) \geq T_2\left(W_\vartheta^0\right) = W_\vartheta^0, \end{cases} \implies B_\vartheta\left(V_\vartheta^0\right) \geq \max\left\{U_\vartheta^0, W_\vartheta^0\right\} = V_\vartheta^0.$$

By iterating the Bellman operator B_ϑ, one obtains $V_\vartheta \geq V_\vartheta^0$, which implies $a_\vartheta \geq 0, b_\vartheta \geq 0$.

It follows from (35) and (36) that U_ϑ and W_ϑ have same values at $x = \bar{x}_\vartheta - 1$ and $\bar{x}_\vartheta$:

$$U_\vartheta(\bar{x}_\vartheta - 1) = W_\vartheta(\bar{x}_\vartheta - 1), \qquad U_\vartheta(\bar{x}_\vartheta) = W_\vartheta(\bar{x}_\vartheta). \tag{37}$$

One also has $\quad T_1(U_\vartheta)(\bar{x}_\vartheta) = U_\vartheta(\bar{x}_\vartheta) = W_\vartheta(\bar{x}_\vartheta) = T_2(W_\vartheta)(\bar{x}_\vartheta) = T_2(U_\vartheta)(\bar{x}_\vartheta),$

where the last equality uses (37). It then follows that

$$U_\vartheta(\bar{x}_\vartheta) - U_\vartheta(\bar{x}_\vartheta + 1) = P.$$

By an abuse of notation, I use U_ϑ and W_ϑ to denote the functions given by (35) and (36) respectively on the entire real line $\mathbb{R}$. There exists some $\tilde{x}_\vartheta \in (\bar{x}_\vartheta, \bar{x}_\vartheta + 1)$ such that

$$U_\vartheta'(\tilde{x}_\vartheta) = -P. \tag{38}$$

Similarly, there exists some $\hat{x}_\vartheta \in (\bar{x}_\vartheta - 1, \bar{x}_\vartheta + 1)$ such that

$$W_\vartheta'(\hat{x}_\vartheta) = -P, \tag{39}$$

Plugging the expressions of U_ϑ and W_ϑ into (37) to (39), one obtains

$$-\frac{2\beta}{r}\tilde{x}_\vartheta + a_\vartheta d_\vartheta \sinh(d_\vartheta \tilde{x}_\vartheta) = -P, \tag{40}$$

$$-\frac{2\beta}{r}\hat{x}_\vartheta + \frac{\vartheta}{r}\frac{\beta}{r} + b_\vartheta c_\vartheta e^{c_\vartheta \hat{x}_\vartheta} = -P. \tag{41}$$

$$\frac{\vartheta}{2r}\left(P - \frac{\beta}{r}\right) + a_\vartheta \cosh(d_\vartheta \bar{x}_\vartheta) = \frac{\vartheta}{r}\frac{\beta}{r}\bar{x}_\vartheta - \left(\frac{\vartheta}{r}\right)^2 \frac{\beta}{2r} + b_\vartheta e^{c_\vartheta \bar{x}_\vartheta}, \tag{42}$$

Equation (41) and $b_\vartheta \geq 0$ imply that

$$0 \leq -\frac{r}{\vartheta}\, b_\vartheta c_\vartheta e^{c_\vartheta \hat{x}_\vartheta} = \frac{\beta}{r}\left(1 - \frac{2r}{\vartheta}\hat{x}_\vartheta\right) + \frac{r}{\vartheta}\, P.$$

Thus, $b_\vartheta e^{c_\vartheta \bar{x}_\vartheta} = O(\vartheta^2)$ and $\hat{x}_\vartheta = O(\vartheta)$. I multiply (41) by $\vartheta/2r$ and subtract by (42),

$$a_\vartheta \cosh(d_\vartheta \bar{x}_\vartheta) = b_\vartheta O(\vartheta^{-1}) e^{c_\vartheta \bar{x}_\vartheta} + O(\vartheta) = O(\vartheta). \tag{43}$$

I show that $\hat{x}_\vartheta = o(\vartheta)$. If this is not the case, then there exists a sequence $(\vartheta_\ell)_{\ell \geq 0}$ going to infinity and $\hat{x}_{\vartheta_\ell} = \Theta(\vartheta_\ell)$ as ℓ goes to infinity. It then follows from (40) that

$$a_{\vartheta_\ell} \sinh(d_{\vartheta_\ell} \tilde{x}_{\vartheta_\ell}) = \Theta\left(\vartheta_\ell^{3/2}\right), \qquad \text{thus} \quad a_{\vartheta_\ell} \cosh(d_{\vartheta_\ell} \tilde{x}_{\vartheta_\ell}) = \Theta\left(\vartheta_\ell^{3/2}\right).$$

This contradicts equation (43). Therefore, $\hat{x} = o(\vartheta)$, and thus $\bar{x}_\vartheta = o(\vartheta)$.

I multiply (41) by $\vartheta/2r$ and subtract by (42), to derive a higher order Taylor expansion

$$\begin{aligned} a_\vartheta \cosh(d_\vartheta \bar{x}_\vartheta) &\sim \frac{\beta}{r}\frac{\vartheta}{r}. \\ \Longrightarrow a_\vartheta d_\vartheta \sinh(d_\vartheta \bar{x}_\vartheta) &\sim \frac{2\beta}{r}\sqrt{\frac{\vartheta}{r}} \tanh(d_\vartheta \bar{x}_\vartheta). \end{aligned} \tag{44}$$

It then follows from (40) that $d_\vartheta \bar{x}_\vartheta \sim \tanh(d_\vartheta \bar{x}_\vartheta)$. However, the equation $y = \tanh y$ does not have non-zero solution. Thus, $\lim_\vartheta d_\vartheta \bar{x}_\vartheta = 0$. A Taylor expansion applied to (44) gives

$$a_\vartheta = \frac{\beta\vartheta}{r^2} - \frac{\beta\vartheta}{2r^2}\, d_\vartheta^2 \bar{x}_\vartheta^2 + O(1). \tag{45}$$

Using equation (45), another Taylor expansion applied to equation (40) leads to

$$-\frac{\beta\vartheta}{3r^2}\, d_\vartheta^4\, \bar{x}_\vartheta^3 = -P,$$

$$\text{which implies} \qquad \tilde{x}_\vartheta = \Theta\left(\vartheta^{1/3}\right), \qquad \bar{x}_\vartheta = \Theta\left(\vartheta^{1/3}\right), \qquad V_\vartheta(0) - V_\vartheta(\bar{x}_\vartheta) = O\left(\vartheta^{1/3}\right). \tag{46}$$

One can further obtain the following expansions, which are useful for proving Theorem 3,

$$V_\vartheta(\bar{x}_\vartheta - 1) - V_\vartheta(\bar{x}_\vartheta) = P - \Theta\left(\vartheta^{-2/3}\right), \quad V_\vartheta(\bar{x}_\vartheta) - V_\vartheta(\bar{x}_\vartheta + 1) = P + \Theta\left(\vartheta^{-2/3}\right). \tag{47}$$

E.3 Proof of Proposition 8

Part (i): For some $m \geq 1$ and $P > 0$, I let $\vartheta = 2\lambda((n-m)\theta_m/m + \theta_{m-1})$. With reparametrization, I write $C(\vartheta)$ for $C(n, \lambda, m)$. The dealer value $V_\vartheta(0)$ increases superlinearly with ϑ (Lemma 10). Thus, the individual dealer inventory cost $C(\vartheta)$ is strictly concave in ϑ. Since $V_\vartheta(0) \leq \vartheta P/r$, then $C(\vartheta) \geq 0$. Hence, it must be that $C(\vartheta)$ is strictly increasing in $\vartheta \in \mathbb{R}^+$.

Equation (45) implies that $a_\vartheta = \beta\vartheta/r^2 + O\left(\vartheta^{2/3}\right)$. Letting $x = 0$ in (35), one has

$$V_\vartheta(0) = \frac{\vartheta P}{r} + O\left(\vartheta^{\frac{2}{3}}\right) \quad \Longrightarrow \quad C(\vartheta) = O\left(\vartheta^{2/3}\right).$$

Part (ii): One has $m\vartheta_m < (m+1)\vartheta_{m+1}$. Since the individual dealer inventory cost $C(\vartheta)$ is strictly increasing and strictly concave in ϑ, it follows from Jensen's inequality that

$$mC(\vartheta_m) = mC(\vartheta_m) + C(0) < (m+1)\, C\left(\frac{m\vartheta_m}{m+1}\right) < (m+1)\, C(\vartheta_{m+1}).$$

E.4 Proof of Proposition 9

Tthe utility $\Phi_{m,P^*(m)}$ of a buyside firm induced by $\sigma^*(m)$ $(m \leq \overline{m})$ is given by

$$\Phi_{m,P^*(m)} = \frac{2\lambda\theta_m(\pi - P^*(m)) - mc}{r}.$$

As n goes to infinity, it follows from (31) that

$$V_{n-m,m,P^*(m)}(0) \sim 2n\lambda\, \frac{\theta_m}{m}\, P^*(m). \tag{48}$$

Thus, $\quad U_m = (n-m)\Phi_{m,P^*(m)} + mV_{n-m,m,P^*(m)}(0)$

$$\sim \left(\frac{2\lambda\theta_m\pi - mc}{r}\right) n = \left[\sum_{1 \leq m' \leq m} (\theta_{m'} - \theta_{m'-1})P^*\left(m'\right)\right] \frac{2\lambda n}{r} \equiv g(m)\, \frac{2\lambda n}{r}.$$

Since $P^*(m') > 0$ for every $1 \leq m' \leq \overline{m}$, then $g(m)$ is strictly increasing in m for $0 \leq m \leq \overline{m}$. The asymptotic equivalence above implies that there exists some integer $n_0 > 0$, if the total number of agent $n > n_0$, the welfare U_m is strictly increasing in m for $0 \leq m \leq \overline{m}$.

F Proofs from Section 5

F.1 Proof of Theorem 3

Fixing an arbitrary $\varepsilon > 0$ and $(\beta, \pi, \lambda, \theta, c, r)$, for n sufficiently large, I first establish that a given dealer's strategy is optimal against $\hat{\sigma}^*_m$, then that a buyside firm's strategy is ε-optimal. The proof is based on $m = 2$ for ease of exposition. The same steps apply to general $m \leq m^*$.

For a given dealer j_1, its expected continuation utility $\widehat{V}_\vartheta(x, y)$, conditional on its information and the inventory y of the other dealer j_2, is a function of the two dealers' current inventories (x, y) only. The function $\widehat{V}_\vartheta$ satisfies

$$
\begin{aligned}
& r\widehat{V}_\vartheta(x, y) \qquad (49) \\
= \quad & \frac{\vartheta}{2}\Big[\, \mathbb{1}_{\{x<\bar{x}_\vartheta\}} \Big(\widehat{V}_\vartheta(x+1, y) + P^*(2) - \widehat{V}_\vartheta(x, y)\Big) + \mathbb{1}_{\{x>-\bar{x}_\vartheta\}} \Big(\widehat{V}_\vartheta(x-1, y) + P^*(2) - \widehat{V}_\vartheta(x, y)\Big)\Big] \\
& + \frac{\vartheta}{2}\Big[\, \mathbb{1}_{\{y<\bar{x}_\vartheta\}} \Big(\widehat{V}_\vartheta(x, y+1) - \widehat{V}_\vartheta(x, y)\Big) + \mathbb{1}_{\{y>-\bar{x}_\vartheta\}} \Big(\widehat{V}_\vartheta(x, y-1) - \widehat{V}_\vartheta(x, y)\Big)\Big] \\
& + \xi\, \Big[\widehat{V}_\vartheta\big(x + \hat{q}(x, y),\ y - \hat{q}(x, y)\big) - \widehat{V}_\vartheta(x, y) - \hat{p}(x, y)\Big] - \beta x^2
\end{aligned}
$$

The interdealer trade terms $(\hat{p}, \hat{q})$ are antisymmetric, that is, $\hat{p}(x, y) = -\hat{p}(y, x)$, $\hat{q}(x, y) = -\hat{q}(y, x)$. Moreover, $(\hat{p}, \hat{q})$ maximize the Nash product:

$$
\begin{aligned}
& \underset{p\in\mathbb{R}, q\in\mathbb{Z}}{\operatorname{argmax}} \Big[\widehat{V}_\vartheta(x+q, y-q) - \widehat{V}_\vartheta(x, y) - p\Big]\Big[\widehat{V}_\vartheta(y-q, x+q) - \widehat{V}_\vartheta(y, x) + p\Big] \\
& \text{subject to } \ \widehat{V}_\vartheta(x+q, y-q) - \widehat{V}_\vartheta(x, y) - p \geq 0, \quad \widehat{V}_\vartheta(y-q, x+q) - \widehat{V}_\vartheta(y, x) + p \geq 0.
\end{aligned} \qquad (50)
$$

Letting $q_1(x, y) = x + \hat{q}(x, y)$ and $q_2(x, y) = y - \hat{q}(x, y)$ be dealers' post-trade inventories, I show that there is a $\widehat{V}$ solving the system above, and a unique one such that $q_1(x, y)$ and $q_2(x, y)$ are weakly increasing in (x, y), which is natural given dealers' risk sharing motives.

Step 1: I show that if there exists a function $\widehat{V}_\vartheta$ that satisfies the HJB equation above, then

$$
\sup_{\substack{-\bar{x}_\vartheta-1\leq x\leq\bar{x}_\vartheta \\ |y|\leq\bar{x}_\vartheta}} |\Delta_1(x, y)| = O(1/\vartheta), \qquad \sup_{\substack{|x|\leq\bar{x}_\vartheta \\ -\bar{x}_\vartheta-1\leq y\leq\bar{x}_\vartheta}} |\Delta_2(x, y)| = O(1/\vartheta), \qquad \text{as } \vartheta \to \infty. \quad (51)
$$

where V_ϑ is the value function of a dealer in the symmetric-agent model, and

$$\Delta_1(x,y) = \left[\widehat{V}_\vartheta(x,y) - \widehat{V}_1(x+1,y)\right] - \left[V_\vartheta(x) - V_\vartheta(x+1)\right], \;\; \Delta_2(x,y) = \widehat{V}_\vartheta(x,y) - \widehat{V}_\vartheta(x,y+1).$$

I let $\widehat{S}_\vartheta(x,y) = \widehat{V}_\vartheta(x,y) + \widehat{V}_\vartheta(y,x)$. It follows from (50) that

$$\widehat{V}_\vartheta\big(q_1(x,y), q_2(x,y)\big) - \widehat{V}_\vartheta(x,y) - \hat{p}(x,y) = \frac{1}{2}\left[\widehat{S}_\vartheta\big(q_1(x,y), \hat{q}_2(x,y)\big) - \widehat{S}_\vartheta(x,y)\right]$$

Then $\widehat{S}_\vartheta$ solves the fixed point problem $f = H_\xi(f)$, where

$$\begin{aligned} H_\xi(f)(x,y) = \frac{1}{r+2\vartheta+\xi}\Bigg(\quad & \frac{\vartheta}{2}\Big[\, \mathbb{1}_{\{x<\bar{x}_\vartheta\}}\big(f(x+1,y) + P^*(2)\big) + \mathbb{1}_{\{x\geq\bar{x}_\vartheta\}}\, f(x,y) + \\ & \qquad \mathbb{1}_{\{x>-\bar{x}_\vartheta\}}\big(f(x-1,y) + P^*(2)\big) + \mathbb{1}_{\{x\leq-\bar{x}_\vartheta\}}\, f(x,y)\Big] \\ + \; & \frac{\vartheta}{2}\Big[\, \mathbb{1}_{\{y<\bar{x}_\vartheta\}}\big(f(x,y+1) + P^*(2)\big) + \mathbb{1}_{\{y\geq\bar{x}_\vartheta\}}\, f(x,y) + \\ & \qquad \mathbb{1}_{\{y>-\bar{x}_\vartheta\}}\big(f(x,y-1) + P^*(2)\big) + \mathbb{1}_{\{y\leq-\bar{x}_\vartheta\}}\, f(x,y)\Big] \\ + \; & \xi\, f(q_1(x,y), q_2(x,y)) - \beta\left(x^2+y^2\right)\Bigg) \end{aligned}$$

Letting $S_\vartheta(x,y) = V_\vartheta(x) + V_\vartheta(y)$, then the function S_ϑ solves the same system as the one above for $\widehat{S}$, with $\xi = 0$ reflecting the absence of an interdealer market. That is, $S_\vartheta = H_0(S_\vartheta)$. The operator H_ξ satisfies the Blackwell's sufficient conditions, thus is a contraction mapping on the set of functions from A_ϑ to $\mathbb{R}$, where $A_\vartheta = [-\bar{x}_\vartheta - 1, -\bar{x}_\vartheta + 1]^2$, with a contraction factor $(2\vartheta+\xi)/(r+2\vartheta+\xi)$. Letting $S^0_\vartheta = S_\vartheta$, and $S^{k+1}_\vartheta = H_\vartheta(S^k_\vartheta)$ for every $k \in \mathbb{N}$, then

$$\sup_{(x,y)\in A_\vartheta}\left|\widehat{S}_\vartheta(x,y) - S_\vartheta(x,y)\right| \leq \sum_{k=0}^{\infty}\left(\frac{2\vartheta+\xi}{r+2\vartheta+\xi}\right)^k \sup_{(x,y)\in A_\vartheta}\left|S^1_\vartheta(x,y) - S^0_\vartheta(x,y)\right|$$

Then (46) implies

$$\begin{aligned} & \max_{(x,y)\in A_\vartheta} S_\vartheta(x,y) - \min_{(x,y)\in A_\vartheta} S_\vartheta(x,y) = O\left(\vartheta^{1/3}\right) \\ \implies & \sup_{(x,y)\in A_\vartheta}\left|S^1_\vartheta(x,y) - S^0_\vartheta(x,y)\right| = O\left(\vartheta^{-2/3}\right) \\ \implies & \sup_{(x,y)\in A_\vartheta}\left|\widehat{S}_\vartheta(x,y) - S_\vartheta(x,y)\right| = O\left(\vartheta^{1/3}\right) \\ \implies & \max_{(x,y)\in A_\vartheta} \widehat{S}_\vartheta(x,y) - \min_{(x,y)\in A_\vartheta} \widehat{S}_\vartheta(x,y) = O\left(\vartheta^{1/3}\right). \end{aligned} \tag{52}$$

Then the value function $\widehat{V}_\vartheta$ solves the fixed point problem $f = B_\xi(f)$, where

$$
\begin{aligned}
B_\xi(f)(x,y) = \frac{1}{r+2\vartheta}\Bigg(\quad & \frac{\vartheta}{2}\Big[\ \mathbb{1}_{\{x<\bar{x}_\vartheta\}}\big(f(x+1,y)+P^*(2)\big) + \mathbb{1}_{\{x\geq\bar{x}_\vartheta\}}\, f(x,y) + \\
& \qquad \mathbb{1}_{\{x>-\bar{x}_\vartheta\}}\big(f(x-1,y)+P^*(2)\big) + \mathbb{1}_{\{x\leq-\bar{x}_\vartheta\}}\, f(x,y)\Big] \\
& + \frac{\vartheta}{2}\Big[\ \mathbb{1}_{\{y<\bar{x}_\vartheta\}}\, f(x,y+1) + \mathbb{1}_{\{y\geq\bar{x}_\vartheta\}}\, f(x,y) + \\
& \qquad \mathbb{1}_{\{y>-\bar{x}_\vartheta\}}\, f(x,y-1) + \mathbb{1}_{\{y\leq-\bar{x}_\vartheta\}}\, f(x,y)\Big] \\
& + \frac{\xi}{2}\Big[\ \widehat{S}_\vartheta(q_1(x,y), q_2(x,y)) - \widehat{S}_\vartheta(x,y)\Big] - \beta x^2\Bigg)
\end{aligned}
$$

The same argument establishing (52) implies that

$$
\sup_{(x,y)\in A_\vartheta}\left|\widehat{V}_\vartheta(x,y) - V_\vartheta(x,y)\right| = O\left(\vartheta^{1/3}\right) \tag{53}
$$

Subtracting $V_\vartheta = B_0\,(V_\vartheta)$ from $\widehat{V}_\vartheta = B_\xi(\widehat{V}_\vartheta)$ gives, $\forall (x,y)\in[-\bar{x}_\vartheta-1,\bar{x}_\vartheta+1]^2\backslash\{\bar{x}_\vartheta,\bar{x}_\vartheta\}$,

$$
\begin{aligned}
&\mathbb{1}_{\{x>-\bar{x}_\vartheta\}}\Delta_1(x-1,y) - \mathbb{1}_{\{x<\bar{x}_\vartheta\}}\Delta_1(x,y) - \mathbb{1}_{\{y<\bar{x}_\vartheta\}}\Delta_2(x,y) + \mathbb{1}_{\{y<-\bar{x}_\vartheta\}}\Delta_2(x,y-1) \\
&= \frac{1}{\vartheta}\left(2r\left[\widehat{V}_\vartheta(x,y) - V_\vartheta(x,y)\right] - \xi\left[\ \widehat{S}_\vartheta\big(x+\hat{q}(x,y),\ y-\hat{q}(x,y)\big) - \widehat{S}_\vartheta(x,y)\right]\right)
\end{aligned} \tag{54}
$$

Some simple algebra also implies, for every $(x,y)\in[-\bar{x}_\vartheta-1,\bar{x}_\vartheta]^2$,

$$
\Delta_1(x,y) + \Delta_2(x+1,y) - \Delta_1(x,y+1) - \Delta_2(x,y) = 0. \tag{55}
$$

Equations (54) and (55) form a fully determined linear system on $\Delta_1(x,y)$ for $(x,y)\in A_{1,\vartheta} := [-\bar{x}_\vartheta-1,\bar{x}_\vartheta]\times[-\bar{x}_\vartheta-1,\bar{x}_\vartheta+1]$, and $\Delta_2(x,y)$ for $(x,y)\in A_{2,\vartheta} := [-\bar{x}_\vartheta-1,\bar{x}_\vartheta+1]\times[-\bar{x}_\vartheta-1,\bar{x}_\vartheta]$. One can write this linear system into a matrix form $M_\vartheta z_\vartheta = w_\vartheta$, where z_ϑ denotes the vector of unknowns. Equations (52) and (53) imply that $||w_\vartheta||_\infty = O\left(\vartheta^{-2/3}\right)$. To establish that $||\Delta_\vartheta||_\infty = O\left(\vartheta^{-2/3}\right)$, it is sufficient to show that $\left\|M_\vartheta^{-1}\right\|_\infty \leq c^{-1}$ for some constant $c>0$, which is equivalent to $||M_\vartheta z||_\infty \geq c$ for every vector z such that $||z||_\infty = 1$. Since $M_\vartheta z$ is linear in every component of z, then maximizing $||M_\vartheta z||_\infty$ implies that every component of z is either 1 or -1. Since all entries of M_ϑ are integers, then the vector $M_\vartheta z$ is also integer-valued if z is. Since M_ϑ is invertible, thus $||M_\vartheta z||_\infty \geq 1$ for every vector z such that $||z||_\infty = 1$. Hence, $||z_\vartheta||_\infty = O\left(\vartheta^{-2/3}\right)$. Furthermore, I will show that $||z_\vartheta||_\infty = O\left(\vartheta^{-1}\right)$. Subtracting (54) for (x, y) and $(x+1, y)$ gives, for every $(x, y) \in [-\bar{x}_\vartheta - 1, \bar{x}_\vartheta] \times [-\bar{x}_\vartheta - 1, \bar{x}_\vartheta + 1] \backslash \{\bar{x}_\vartheta, \bar{x}_\vartheta\}$,

$$
\begin{aligned}
&\mathbb{1}_{\{x>-\bar{x}_\vartheta\}}\Delta_1(x-1,y) - \mathbb{1}_{\{x>-\bar{x}_\vartheta-1\}}\Delta_1(x,y) - \mathbb{1}_{\{x<\bar{x}_\vartheta\}}\Delta_1(x,y) + \mathbb{1}_{\{x<\bar{x}_\vartheta-1\}}\Delta_1(x+1,y) \\
&- \mathbb{1}_{\{y<\bar{x}_\vartheta\}}[\Delta_2(x,y) - \Delta_2(x+1,y)] + \mathbb{1}_{\{y<-\bar{x}_\vartheta\}}[\Delta_2(x,y-1) - \Delta_2(x+1,y-1)] \\
= &\frac{1}{\vartheta}\Big[2r\Delta_1(x,y) - \xi\Big(\quad \Big[\widehat{S}_\vartheta(q_1(x,y), q_2(x,y)) - \widehat{S}_\vartheta(q_1(x+1,y), q_2(x+1,y))\Big] \\
&\qquad\qquad - \Big[\widehat{S}_\vartheta(x,y) - \widehat{S}_\vartheta(x+1,y)\Big]\Big)\Big]
\end{aligned}
$$

Since $\Delta_2(x,y) - \Delta_2(x+1,y) = \Delta_1(x,y) - \Delta_1(x,y+1)$ and $\Delta_2(x,y-1) - \Delta_2(x+1,y-1) = \Delta_1(x,y-1) - \Delta_1(x,y)$, the left hand side of the equation above can be written as a function of only the Δ_1 terms. Therefore, the linear system above, fully determined, can be written into a matrix form $\widetilde{M}_\vartheta\, z_{1,\vartheta} = w_{1,\vartheta}$, where $z_{1,\vartheta}$ denotes the vector $\Delta_1(x,y)$, for $(x,y) \in A_{1,\vartheta}$. To bound $||z_{1,\vartheta}||_\infty$, it is sufficient to bound $||w_{1,\vartheta}||_\infty$. Since $q_k(x,y)$ $(k = 1, 2)$ is weakly increasing in (x,y), and $q_1(x,y) + q_2(x,y) = x + y$, then $0 \leq q_k(x+1,y) \leq 1$. Thus,

$$
\sup_{(x,y)\in A_{1,\vartheta}} \Big|\widehat{S}_\vartheta(q_1(x,y), q_2(x,y)) - \widehat{S}_\vartheta(q_1(x+1,y), q_2(x+1,y))\Big| = O(1).
$$

Therefore, $||w_{1,\vartheta}||_\infty = O(\vartheta^{-1})$. Since $\widetilde{M}_\vartheta$ is an invertible matrix with integer entries, one has $||z_{1,\vartheta}||_\infty = O(\vartheta^{-1})$. Similarly, $||z_{2,\vartheta}||_\infty = O(\vartheta^{-1})$, where $z_{2,\vartheta}$ denotes the vector of $\Delta_2(x,y)$, for $(x,y) \in A_{2,\vartheta}$. This establishes (51), completing *Step 1.*

Step 2: I show that there exists a unique function $\widehat{V}_\vartheta$ solving (49) subject to (50). It follows from (47) and (51) that if ϑ is sufficiently large, then for every $(x,y) \in [-\bar{x}_\vartheta, \bar{x}_\vartheta]^2$,

$$
\begin{aligned}
\mathbb{1}_{\{x<\bar{x}_\vartheta\}}\Big(\widehat{S}(x+1,y) + P^*(2)\Big) + \mathbb{1}_{\{x\geq\bar{x}_\vartheta\}}\,\widehat{S}(x,y) &= \Big(\widehat{S}(x+1,y) + P^*(2)\Big) \vee \widehat{S}(x,y), \\
\mathbb{1}_{\{x>-\bar{x}_\vartheta\}}\Big(\widehat{S}(x-1,y) + P^*(2)\Big) + \mathbb{1}_{\{x\leq-\bar{x}_\vartheta\}}\,\widehat{S}(x,y) &= \Big(\widehat{S}(x-1,y) + P^*(2)\Big) \vee \widehat{S}(x,y), \\
\mathbb{1}_{\{y<\bar{x}_\vartheta\}}\Big(\widehat{S}(x,y+1) + P^*(2)\Big) + \mathbb{1}_{\{y\geq\bar{x}_\vartheta\}}\,\widehat{S}(x,y) &= \Big(\widehat{S}(x,y+1) + P^*(2)\Big) \vee \widehat{S}(x,y), \\
\mathbb{1}_{\{y>-\bar{x}_\vartheta\}}\Big(\widehat{S}(x,y-1) + P^*(2)\Big) + \mathbb{1}_{\{y\leq-\bar{x}_\vartheta\}}\,\widehat{S}(x,y) &= \Big(\widehat{S}(x-1,y) + P^*(2)\Big) \vee \widehat{S}(x,y).
\end{aligned}
$$

Hence, the function $\widehat{S}_\vartheta$ is the unique solution to the fixed point problem $f = \widehat{H}_\xi(f)$, where

$$
\begin{aligned}
&\widehat{H}_\xi(f)(x,y) \\
&= \frac{1}{r+2\vartheta+\xi}\Bigg(\quad \frac{\vartheta}{2}\Big[\,\big(f(x+1,y)+P^*(2)\big)\vee f(x,y) + \big(f(x-1,y)+P^*(2)\big)\vee f(x,y)\Big] \\
&\qquad\qquad + \frac{\vartheta}{2}\Big[\,\big(f(x,y+1)+P^*(2)\big)\vee f(x,y) + \big(f(x,y-1)+P^*(2)\big)\vee f(x,y)\Big] \\
&\qquad\qquad + \ \xi\, f(q_1(x,y), q_2(x,y)) - \beta\left(x^2+y^2\right)\Bigg)
\end{aligned}
$$

The operator $\widehat{H}_\vartheta$ is contracting and preserves symmetry and concavity, function $\widehat{S}_\vartheta(\cdot,\cdot)$ is thus strictly concave and symmetric. Maximizing $\widehat{S}_\vartheta(q_1,q_2)$ subject to $q_1+q_2 = x+y$ gives

$$
q_1(x,y) = \left\lfloor \frac{x+y}{2} \right\rfloor, \qquad q_2(x,y) = \left\lceil \frac{x+y}{2} \right\rceil.
$$

Hence, the function $\widehat{V}$ is uniquely determined by the linear system $\widehat{V}_\vartheta = B_\xi(\widehat{V}_\vartheta)$. Conversely, if one lets $\widehat{S}_\vartheta$ be the unique solution to the fixed point problem $\widehat{H}_\xi(f) = f$, and lets $\widehat{V}_\vartheta$ be determined by $\widehat{V}_\vartheta = B_\xi(\widehat{V}_\vartheta)$, then $\widehat{V}_\vartheta$ solves the system (49) subject to (50).

Step 3: I show that the strategy of a dealer j_1 is optimal against $\hat{\sigma}_2^*$. If ϑ is sufficiently large,

$$
\widehat{V}_\vartheta(\bar{x}_\vartheta - 1, y) - \widehat{V}_\vartheta(\bar{x}_\vartheta, y) < P^*(2), \quad \widehat{V}_\vartheta(\bar{x}_\vartheta) - \widehat{V}_\vartheta(\bar{x}_\vartheta+1, y) > P^*(2).
$$

for every $y \in [-\bar{x}_\vartheta, \bar{x}_\vartheta]$ (this follows from (47) and (51)). The same argument as in *Step 2* implies that the function $\widehat{V}_\vartheta(x,y)$ is concave in $x \in \mathbb{R}$ for any fixed $y \in [-\bar{x}_\vartheta, \bar{x}_\vartheta]$. Hence,

$$
\widehat{V}_\vartheta(x,y) - \widehat{V}_\vartheta(x+1,y) \begin{cases} < P^*(2) & \text{if } x < \bar{x}_\vartheta, \\ > P^*(2) & \text{if } x \leq \bar{x}_\vartheta. \end{cases}
$$

for every $y \in [-\bar{x}_\vartheta, \bar{x}_\vartheta]$. That is, $\hat{p}_m$ is optimal for j_1 if j_1 were restricted to quotes $\pm P^*(2)$ and $\pm\infty$. It remains to show that j_1 has no incentive to quote other prices. I let

$$
\vartheta_0 = 2\lambda(n-m)\theta_m/m, \quad \vartheta_{-1} = 2\lambda(n-m-1)\theta_m x/m, \quad D_\vartheta = [-\bar{x}_\vartheta, \bar{x}_\vartheta]^2, \quad D_0 = [-\bar{x}_{\vartheta_0}, \bar{x}_{\vartheta_0}]^2.
$$

To show that j_1 has no incentive to gouge is equivalent to show that

$$\pi - P^*(2) < \min_{(x,y)\in D_0} \left[\widehat{V}_{\vartheta_0}(x,y) - \widehat{V}_{\vartheta_{-1}}(x,y)\right]. \tag{56}$$

I formally differentiate (49) with respect to ϑ to obtain $\frac{\partial}{\partial\vartheta}\widehat{V}_\vartheta = \widehat{T}_\vartheta\left(\frac{\partial}{\partial\vartheta}\widehat{V}_\vartheta\right)$, where

$$\begin{aligned}
\widehat{T}_\vartheta(f)(x,y) = \frac{1}{r+2\vartheta+\xi}\Bigg(\quad & \frac{\vartheta}{2}\Big[\mathbb{1}_{\{x<\bar{x}_\vartheta\}} f(x+1,y) + \mathbb{1}_{\{x\geq\bar{x}_\vartheta\}} f(x,y) + \\
& \quad \mathbb{1}_{\{x>-\bar{x}_\vartheta\}} f(x-1,y) + \mathbb{1}_{\{x\leq-\bar{x}_\vartheta\}} f(x,y)\Big] \\
& + \frac{\vartheta}{2}\Big[\mathbb{1}_{\{y<\bar{x}_\vartheta\}} f(x,y+1) + \mathbb{1}_{\{y\geq\bar{x}_\vartheta\}} f(x,y) + \\
& \quad \mathbb{1}_{\{y>-\bar{x}_\vartheta\}} f(x,y-1) + \mathbb{1}_{\{y\leq-\bar{x}_\vartheta\}} f(x,y)\Big] \\
& + \xi f\left(\left\lfloor\frac{x+y}{2}\right\rfloor, \left\lceil\frac{x+y}{2}\right\rceil\right) + \vartheta\psi(x) + \delta(x,y)\Bigg),
\end{aligned}$$

$$\delta(x,y) = \frac{1}{2}\left[\mathbb{1}_{\{x>-\bar{x}_\vartheta\}}\Delta_1(x-1,y) - \mathbb{1}_{\{x<\bar{x}_\vartheta\}}\Delta_1(x,y) - \mathbb{1}_{\{y<\bar{x}_\vartheta\}}\Delta_2(x,y) + \mathbb{1}_{\{y<-\bar{x}_\vartheta\}}\Delta_2(x,y-1)\right],$$

and ψ is given by (13). It follows from (51), (16) and $\bar{x} = O\left(\vartheta^{1/3}\right)$ that as $\vartheta \to \infty$,

$$\begin{aligned}
& \sup_{(x,y)\in D_\vartheta} |\delta(x,y)| = O(1/\vartheta), \\
& \sup_{(x,y)\in D_\vartheta} \left|\frac{\partial}{\partial\vartheta} V_\vartheta\left(\left\lfloor\frac{x+y}{2}\right\rfloor\right) - \frac{\partial}{\partial\vartheta} V_\vartheta(x)\right| = O\left(\vartheta^{-2/3}\right), \\
\Longrightarrow \quad & \sup_{(x,y)\in D_\vartheta} \left|\widehat{T}_\vartheta\left(\frac{\partial}{\partial\vartheta} V_\vartheta\right)(x,y) - \frac{\partial}{\partial\vartheta} V_\vartheta(x)\right| = O\left(\vartheta^{-5/3}\right)
\end{aligned}$$

Since the operator $\widehat{T}_\vartheta$ is contracting with a contraction factor $(2\vartheta+\xi)/(r+2\vartheta+\xi)$, one has

$$\sup_{(x,y)\in D_\vartheta} \left|\frac{\partial}{\partial\vartheta}\widehat{V}_\vartheta(x,y) - \frac{\partial}{\partial\vartheta} V_\vartheta(x)\right| = O\left(\vartheta^{-2/3}\right).$$

$$\begin{aligned}
\text{then } \min_{(x,y)\in D_0}\left[\widehat{V}_{\vartheta_0}(x,y) - \widehat{V}_{\vartheta_{-1}}(x,y)\right] &= \min_{(x,y)\in D_0}\left[\int_{\vartheta_{-1}}^{\vartheta_0} \frac{\partial}{\partial\vartheta}\widehat{V}_\vartheta(x,y)d\vartheta\right] \\
= \min_{|x|\leq\bar{x}_{\vartheta_0}}\left[\int_{\vartheta_{-1}}^{\vartheta_0}\frac{\partial}{\partial\vartheta} V_\vartheta(x)d\vartheta\right] + O\left(\vartheta^{-2/3}\right) &= \min_{|x|\leq\bar{x}_{\vartheta_0}}\left[V_{\vartheta_0}(x) - V_{\vartheta_{-1}}(x)\right] + O\left(\vartheta^{-2/3}\right).
\end{aligned}$$

Since it is assumed that $m^* \geq 2$, one has

$$\pi - P^*(2) < \min_{|x| \leq \bar{x}_{\vartheta_0}} \left[V_{\vartheta_0}(x) - V_{\vartheta_{-1}}(x) \right],$$

When ϑ_0 is sufficiently large, the desired inequality (56) thus follows.

Hence, the quoting strategy $\hat{p}_m^*$ is optimal for a given dealer j_1 if j_1 knows the inventory of the other dealer j_2 at *all* times. When employing $\hat{p}_m^*$, j_1 does not use this extra inventory information of j_2. For j_1, $\hat{p}_m^*$ is thus optimal against $\hat{\sigma}_2^*$ given its available information.

I next show that the strategy of a buyside firm i is ε-optimal. The intuition is as follows: Under the strategy profile $\hat{\sigma}_m^*$, i receives a trading gain of either $\pi - P^*(m)$ or 0 each time it requests a quote from a dealer counterparty. This payoff is 0 when the the quoting dealer's inventory is on the boundary $\pm\bar{x}_\vartheta$, and the desired trade direction of the RFQ would further expand the this inventory. In this case, the dealer rejects the RFQ by posting either an offer price ∞ or a bid price $-\infty$. When ϑ is large, the probability of this event is arbitrarily close to 0. Therefore, the continuation payoff of i is arbitrarily close to $\Phi_{m,P^*(m)}$. Similarly, his continuation payoff is arbitrarily close to $\Phi_{d,P^*(m)}$ if i has d dealer accounts. Therefore, maintaining m dealer accounts is an ε-optimal strategy for the buyside firm.

Formally, I first consider the benchmark case in which the buyside firm's RFQ is never rejected. Then the maximum attainable continuation utility of i is $\Phi^*_{m,P^*(m)}$. I let $\mathcal{G}(G(m), i)$ be the set of networks in which all agents other than i are connected to each other as in $G(m)$, and $D_\vartheta := [-\bar{x}_\vartheta, \bar{x}_\vartheta]^m$ be the set of all possible dealer inventories in the actual trading game. To distinguish from the symmetric-agent model, I let $\widehat{X}_{jt}$ denote the inventory of agent j at time t. Since the strategy profile $\hat{\sigma}^*(m)$ is stationary, the information available to i at time t can be equivalently represented as an inference distribution of the current state $(G_t, \widehat{X}_t)$. Any such inference distribution – obtained through the Bayes rule whenever possible – assigns probability 1 to the set $\mathcal{G}(G(m), i)$ of networks and to the set D_ϑ of dealer inventories. I let H_t include the support of all such inference distributions. That is,

$$H_t = \{G_t \in \mathcal{G}(G(m), i), \widehat{X}_{jt} \in D_\vartheta\}$$

For every game history $h_t \in H_t$ up to but excluding time t, I let $\Phi_{\hat{\sigma}^*(m)}(h_t)$ be the continuation utility of i after h_t has been realized at time t. In other words, $\Phi_{\hat{\sigma}^*(m)}(h_t)$ is the continuation utility of i if i were able to observe the complete game history h_t. It suffices to show that $\Phi_{\hat{\sigma}^*(m)}(h_t)$ is ε-close to $\Phi^*_{m,P^*(m)}$ for every t and $h_t \in H_t$.

The vector $(\widehat{X}_{Jt})_{t\geq 0}$ of all dealer inventories is a Markov process with state space D_ϑ. I let ∂D_ϑ denote the boundary of D_ϑ, τ_k be the k'th time that either *(i)* a dealer receives a RFQ, or *(ii)* two dealers meet in the interdealer market. Then $(\tau_k)_{k\geq 1}$ are the event times of a Poisson process with intensity $\chi = 2(n-m)\lambda\theta_m + m(m-1)\xi/2$. I let $Z_k \in \{0,1\}$ be the binary variable indicating whether buyside firm i submits a RFQ at τ_k, and j_k be the dealer who receives this RFQ. I let x_{Jt} be the vector of current dealer inventories under h_t, then

$$\begin{aligned}
\Phi^*_{m,P^*(m)} - \Phi_{\hat{\sigma}^*(m)}(h_t) &\leq \mathrm{E}\left(\sum_{\tau_k\geq t} e^{-r(\tau_k-t)}[\pi - P^*(m)]\,\mathbb{1}\{Z_k=1\}\,\mathbb{1}\left\{\left|\widehat{X}_{j_k\tau_k}\right| = \bar{x}_\vartheta\right\}\right) \\
&\leq [\pi - P^*(m)]\,\frac{2\lambda\theta_m}{\chi}\sum_{k\geq 1}\mathrm{E}\big(e^{-r\tilde{\tau}_k}\big)\;\mathrm{P}\Big(\widehat{Y}_k \in \partial D_\vartheta \,|\, \widehat{Y}_0 = x_{Jt}\Big) \\
&= [\pi - P^*(m)]\,\frac{2\lambda\theta_m}{\chi}\sum_{k\geq 1}\left(\frac{\chi}{\chi+r}\right)^k \mathrm{P}\Big(\widehat{Y}_k \in \partial D_\vartheta \,|\, \widehat{Y}_0 = x_{Jt}\Big),
\end{aligned}$$

where $(\widehat{Y}_k)_{k\geq 0}$ is the embedded discrete-time Markov Chain of $(\widehat{X}_{Js})_{s\geq t}$. The second inequality above uses the independence between $(\tau_k)_{k\geq 1}$ and $(\widehat{X}_{J\tau_k})_{k\geq 1}$, between Z_k and $(\tau_k, \widehat{X}_{j_k\tau_k})_{k\geq 1}$, and $\mathrm{P}(Z_k = 1) = 2\lambda\theta_m/\chi$. Lemmas 12 and 13 imply that for every ε, there exists n_0 and $k(n) = O(n^{2/3})$ such that for every $n > n_0$, $k > k(n)$ and $x_{Jt} \in D_\vartheta$,

$$\mathrm{P}(Y_k \in \partial D_\vartheta \,|\, Y_0 = x_{Jt}) < \varepsilon.$$

Hence, there exists some n_1 such that for every $n > n_1$, t and $h_t \in H_t$,

$$\Phi^*_{m,P^*(m)} - \Phi_{\hat{\sigma}^*(m)}(h_t) < 2\lambda\theta_m[\pi - P^*(m)]\left(\frac{k(n)}{\chi} + \frac{\varepsilon}{r}\right) < 4\lambda\theta_m[\pi - P^*(m)]\,\frac{\varepsilon}{r}.$$

Lemma 12. *I let μ denote the stationary distribution of $\widehat{X}_{Jt}$, and $\hat{\mu}$ that of $\widehat{X}_{Jt}$. Then*

$$\hat{\mu}(\partial D_\vartheta) \leq \mu(\partial D_\vartheta)$$

Proof. I show that if $X \sim \mu$ and $\widehat{X} \sim \hat{\mu}$, then

$$\max_{j \in J} |X_j| \overset{d}{\geq} \max_{j \in J} |\widehat{X}_j|, \tag{57}$$

where $\overset{d}{\geq}$ denotes stochastic dominance. It would then follow that

$$\mu(\partial D_\vartheta) = \mathrm{P}\left(\max_{j \in J} |\widehat{X}_j| = \bar{x}_\vartheta\right) \geq \mathrm{P}\left(\max_{j \in J} |X_j| = \bar{x}_\vartheta\right) = \hat{\mu}(\partial D_\vartheta)$$

I let (Y_k) be the embedded discrete-time Markov Chain of (X_{Jt}), and $(y_k, \hat{y}_k, y_{k+1}, \hat{y}_{k+1})$ be integers within $[0, \bar{x}_\vartheta]$ such that $y_k \geq \hat{y}_k$ and $y_{k+1} \geq \hat{y}_{k+1}$. Since interdealer trading can only reduce the maximum dealer inventory, one has

$$\mathrm{P}\left(\max_{j \in J} |Y_{j(k+1)}| \geq y_{k+1} \;\middle|\; \max_{j \in J} |Y_{jk}| = y_k\right) \geq \mathrm{P}\left(\max_{j \in J} |\widehat{Y}_{j(k+1)}| \geq \hat{y}_{k+1} \;\middle|\; \max_{j \in J} |\widehat{Y}_{jk}| = \hat{y}_k\right).$$

Hence, the desired stochastic dominance (57) follows from induction. □

Lemma 13. *As $n \to \infty$, the mixing time of $(\widehat{X}_{Jt})_{t \geq 0}$ is asymptotically bounded by $n^{-1/3}$.*

Proof. I use the *coupling* technique.[18] I consider a lazy version of $\widehat{Y}_k$, which remains in its current position with probability $1/2$ and otherwise moves with the same transition probabilities as $\widehat{Y}_k$. I construct a coupling (Y_k, Z_k) of two lazy chains on D_ϑ, starting from $Y_0 = y$ and $Z_0 = z$ respectively. At each move, either two randomly chosen dealers j_1, j_2 trade in the interdealer market with probability ξ/χ, or one dealer j receives a RFQ. In the first case, a fair coin is tossed to determine which of the two chains (Y_k) or (Z_k) moves. In the second case, if $Y_{jk} \neq Z_{jk}$, then a fair coin is tossed to determine whether (Y_{jk}) or (Z_{jk}) receives the RFQ. If $Y_{jk} = Z_{jk}$, then a fair coin is tossed to determine whether both (Y_{jk}) and (Z_{jk}) receive the RFQ, or none does. Once the two chains Y_{jk} and Z_{jk} collide, thereafter they make identical moves. I let $L_k = Y_k - Z_k$ and $\hat{\mu}_{k,y}$ be the distribution of Y_k. Then

$$\max_{y \in D_\vartheta} ||\hat{\mu}_{k,y} - \hat{\mu}||_{\mathrm{TV}} \leq \max_{y,z \in D_\vartheta} \mathrm{P}_{y,z}(L_k \neq 0). \tag{58}$$

[18] Chapter 5 of Levin, Peres, and Wilmer (2009) provides relevant background for the coupling technique.

However, the process (L_k) is not Markovian. I construct a conditional Markov chain $(\bar{L}_k)$ such that $|\bar{L}_{jk}| \geq |L_{jk}|$ almost surely for every $j \in J$ and $k \geq 0$. The increment $\bar{L}_{k+1} - \bar{L}_k$ is equal to $L_{k+1} - L_k$ unless at the k'th move, *(i)* either Y_{jk} or Z_{jk} (but not both) receives a RFQ, *(ii)* if Y_{jk} receives the RFQ, then $Y_{jk} = \pm\bar{x}_\vartheta$ and the RFQ would further expand Y_{jk}. Likewise if Z_{jk} receives the RFQ. If these two conditions hold, then $L_{k+1} - L_k = 0$ by definition. I let $\bar{L}_{j(k+1)} - \bar{L}_{jk} = \operatorname{sgn} Y_{jk} - \operatorname{sgn} Z_{jk}$. In other words, when Y_{jk} or Z_{jk} receives a RFQ that would make the inventory move beyond the boundary $\pm\bar{x}_\vartheta$, the increment $\bar{L}_{j(k+1)} - \bar{L}_{jk}$ is determined as if the inventory did move beyond the boundary. One can verify that $|\bar{L}_{jk}| \geq |L_{jk}|$ almost surely for every (j,k). I let E_h be the event that no interdealer trade occurs up to period h. Conditional on the event E_h, each $(\bar{L}_{jk})$ for some $j \in J$ is a lazy random walk in $[-2\bar{x}_\vartheta, 2\bar{x}_\vartheta]$ up to period h unless being absorbed by 0. If $\bar{L}_{jk} \neq 0$, it remains at its current position with probability $(m-1)/m$ and moves as a bounded random walk that loops at the end points $\pm 2\bar{x}_\vartheta$ with probability $1/m$. Therefore,

$$\mathrm{P}\big(\bar{L}_k = 0\big) > \mathrm{P}\big(\bar{L}_k = 0 \,|\, E_k\big)\,\mathrm{P}(E_k) = \mathrm{P}^m\left(\bar{L}_{jk} = 0 \,|\, E_k\right)\left(\frac{2(n-m)\lambda\theta_m}{\chi}\right)^k \tag{59}$$

To calculate $\mathrm{P}\big(\bar{L}_k = 0 \,|\, E_k\big)$, I consider the lazy random walk $(\widetilde{L}_k)$ that starts from the same state $y_j - z_j$ as $(\bar{L}_{jk})$ and is not absorbed by 0. Then

$$\mathrm{P}\big(\bar{L}_{jk} = 0 \,|\, E_k\big) = \mathrm{P}\Big(\widetilde{L}_k = 0 \;\Big|\; \widetilde{L}_0 = y_j - z_j\Big).$$

I let $\tau = \min\{k : \widetilde{L}_k = 0\}$ and $f_\ell = \mathrm{E}\Big(\tau \;\Big|\; \widetilde{L}_0 = \ell\Big)$, then $f_0 = 0$ and

$$\begin{aligned}
f_\ell &= \frac{m-1}{m}(1+f_\ell) + \frac{1}{2m}(1+f_{\ell-1}) + \frac{1}{2m}(1+f_{\ell+1}), \qquad 0 < |\ell| < 2\bar{x}_\vartheta,\\
f_{2\bar{x}_\vartheta} &= \frac{m-1}{m}(1+f_{2\bar{x}_\vartheta}) + \frac{1}{2m}(1+f_{2\bar{x}_\vartheta-1}) + \frac{1}{2m}(1+f_{2\bar{x}_\vartheta}).
\end{aligned}$$

One can solve the system above to obtain $f_\ell = m|\ell|\,(4\bar{x}_\vartheta - |\ell| - 1) \leq 2m\bar{x}_\vartheta(2\bar{x}_\vartheta - 1)$. Hence,

$$= \mathrm{P}\Big(\tau > k \;\big|\; \widetilde{L}_0 = y_j - z_j\Big) < \frac{\mathrm{E}\Big(\tau \;\big|\; \widetilde{L}_0 = y_j - z_j\Big)}{k} \leq \frac{2m\bar{x}_\vartheta(2\bar{x}_\vartheta - 1)}{k}$$

It then follows from (58) and (59) that

$$\max_{y \in D_\vartheta} ||\hat{\mu}_{k,y} - \hat{\mu}||_{\mathrm{TV}} \le \mathrm{P}\big(\bar{L}_k \neq 0\big) < 1 - \left(1 - \frac{2m\bar{x}_\vartheta(2\bar{x}_\vartheta - 1)}{k}\right)^m \left(\frac{2(n-m)\lambda\theta_m}{\chi}\right)^k$$

The right hand side is arbitrarily close to 0 if k is appropriately chosen on the order of $O(n^{2/3})$. Hence, the mixing time of $(\widehat{X}_{Jt})$ is asymptotically bounded by $n^{-1/3}$. □

F.2 Proof of Proposition 10

I define the long-run averages of total trade volume and volume in the interdealer market by

$$\mathrm{Vol} = \lim_{T\to\infty} \frac{1}{T} \sum_{i,j \in N} \mathrm{Vol}_{i,j}(T), \qquad \mathrm{Vol}_{\mathrm{ID}} = \lim_{T\to\infty} \frac{1}{T} \sum_{j,j' \in J} \mathrm{Vol}_{j,j'}(T),$$

where $\mathrm{Vol}_{i,j}(T)$ is the total volume traded between agents i and j in the time interval $[0, T]$. The Ergodic Theorem implies that

$$\mathrm{Vol}_{\mathrm{ID}} = m(m-1)\xi\, \mathrm{E}(|\hat{q}(X_j, X_{j'})|)\,, \qquad \text{where } |\hat{q}(X_j, X_{j'})| = \left|\frac{X_j - X_{j'}}{2}\right|.$$

The expectation E is taken with respect to the stationary distribution of X_J. As $n\lambda \to \infty$,

$$\mathrm{E}(|\hat{q}(X_j, X_{j'})|) = \Theta\left((n\lambda)^{\frac{1}{3}}\right), \qquad \text{hence,} \quad \mathrm{Vol}_{\mathrm{ID}} = \Theta\left((n\lambda)^{\frac{1}{3}}\right).$$

Similarly, $\mathrm{Vol} = \Theta(n\lambda)$ as $n \to \infty$. Therefore, $\mathrm{Vol}_{\mathrm{ID}}/\mathrm{Vol} = \Theta\left((n\lambda)^{-2/3}\right)$.

F.3 Proof of Proposition 11

It follows from (48) and (53) that $\widehat{V}_{n-m,m,P^*(m)}(0) \sim V_{n-m,m,P^*(m)}(0)$ as $n \to \infty$. Hence, for every $m \le \overline{m}$, $\widehat{U}_m \sim U_m$ as $n\lambda \to \infty$. Proposition 11 thus follows.

References

Abad, J., I. Aldasoro, C. Aymanns, M. D'errico, L. Fache, R. P. Hoffmann, S. Langfield, M. Neychev, and T. Roukny, 2016, "Shedding light on dark markets: First insights from the new EU-wide OTC derivatives dataset," *ESRB Occasional Paper Series*.

Adrian, T., E. Etula, and H. S. Shin, 2010, "Risk Appetite and Exchange Rates," *NY Fed*.

Adrian, T., M. Fleming, J. Goldberg, M. Lewis, F. Natalucci, and J. Wu, 2013, "Dealer Balance Sheet Capacity and Market Liquidity during the 2013 Selloff in Fixed Income Markets," *FEDS Notes*.

Adrian, T., E. Moench, and H. S. Shin, 2010, "Macro Risk Premium and Intermediary Balance Sheet Quantities," *IMF Economic Review*, 58(1), 179–207.

Afonso, G., A. Kovner, and A. Schoar, 2014, "Trading partners in the interbank lending market," working paper, Federal Reserve Bank of New York and MIT Sloan.

Allen, L., and A. Saunders, 1986, "The large-small bank dichotomy in the federal funds market," *Journal of Banking & Finance*, 10(2), 219–230.

Anderson, R. M., 2000, *Infinitesimal Methods in Mathematical Economics*.

Babus, A., 2016, "The formation of financial networks," *The RAND Journal of Economics*, 47(2), 239–272.

Babus, A., and T.-W. Hu, 2016, "Endogenous Intermediation in Over-the-Counter Markets," *Journal of Financial Economics (Forthcoming)*.

Babus, A., and C. Parlatore, 2016, "Strategic Fragmented Markets," working paper, Federal Reserve Bank of Chicago and NYU Stern.

Bagwell, K., and G. Ramey, 1992, "The Diamond Paradox: A Dynamic Resolution," working paper, Northwestern University and U.C. San Diego.

Bech, M. L., and E. Atalay, 2010, "The topology of the federal funds market," *Physica A: Statistical Mechanics and its Applications*, 389(22), 5223–5246.

Bernheim, B., B. Peleg, and M. D. Whinston, 1987, "Coalition-Proof Nash Equilibria I. Concepts," *Journal of Economic Theory*, 42(1), 1–12.

Blasques, F., F. Bräuning, and I. van Lelyveld, 2015, "A Dynamic Network Model of the Unsecured Interbank Lending Market," working paper, University Amsterdam and BIS.

Boss, M., H. Elsinger, M. Summer, and S. Thurner, 2004, "Network topology of the interbank market," *Quantitative Finance*, 4(6), 677–684.

Brogaard, J., 2010, "High frequency trading and its impact on market quality," working paper, Kellogg School of Management, Northwestern University.

Brogaard, J., T. Hendershott, and R. Riordan, 2014, "High-Frequency Trading and Price Discovery," *Review of Financial Studies*, 27(8), 2267–2306.

Brunnermeier, M. K., and L. H. Pedersen, 2008, "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22(6), 2201–2238.

Cetorelli, N., B. Hirtle, D. P. Morgan, S. Peristiani, and J. Santos, 2007, "Trends in financial market concentration and their implications for market stability," *Economic Policy Review*, 13(Mar), 33–51.

Chang, B., and S. Zhang, 2016, "Endogenous Market Making and Network Formation," working paper, University of Wisconsin and London School of Economics.

Chang, E. J., E. J. A. Lima, S. M. Guerra, and B. M. Tabak, 2008, "Measures of Interbank Market Structure: An Application to Brazil," *Brazilian Review of Econometrics*, 28(2), 163–190.

Colliard, J.-E., T. Foucault, and P. Hoffmann, 2016, "Trading Frictions in the Interbank Market, and the Central Bank," working paper, HEC Paris and European Central Bank.

Comerton-Forde, C., T. Hendershott, C. Jones, P. Moulton, and M. Seasholes, 2010, "Time Variation in Liquidity: The Role of Market-Maker Inventories and Revenues," *The Journal of Finance*, 65(1), 295–331.

Craig, B., and G. von Peter, 2014, "Interbank tiering and money center banks," *Journal of Financial Intermediation*, 23(3), 322–347.

de Roure, C., and C. Wang, 2016, "How Decentralized is the German Bund OTC Market," working paper, Frankfurt School of Finance & Management, Stanford University.

DeMarzo, P. M., M. J. Fishman, and K. M. Hagerty, 2005, "Self-Regulation and Government Oversight," *Review of Economic Studies*, 72(3), 687–706.

Di Maggio, M., A. Kermani, and Z. Song, 2015, "The Value of Trading Relationships in Turbulent Times," working paper, Harvard Business School, U.C. Berkeley and Carey Business School.

Dick-Nielsen, J., and M. Rossi, 2016, "The Cost of Immediacy for Corporate Bonds," working paper, Copenhagen Business School, Texas A&M.

Du, S., and H. Zhu, 2014, "Welfare and Optimal Trading Frequency in Dynamic Double Auctions," working paper, Simon Fraser University and MIT Sloan School of Management.

Duffie, D., 2012, "Market Making Under the Proposed Volcker Rule," working paper, Stanford Graduate School of Business.

Farboodi, M., 2015, "Intermediation and Voluntary Exposure to Counterparty Risk," working paper, Princeton University.

Farboodi, M., G. Jarosch, and R. Shimer, 2016, "Meeting Technologies in Decentralized Asset Markets," working paper, Princeton, Stanford and University of Chicago.

Foley-Fisher, N., S. Gissler, and C. Wang, 2016, "Dealer Inventory and Liquidity in Corporate Bond Markets," working paper, Federal Reserve Board of Governors, Stanford University.

Fricke, D., and T. Lux, 2014, “Core–Periphery Structure in the Overnight Money Market: Evidence from the e-MID Trading Platform,” *Computational Economics*, 45(3), 359–395.

Friewald, N., and F. Nagler, 2016, “Dealer Inventory and the Cross-Section of Corporate Bond Returns,” working paper, Norwegian School of Economics, Bocconi University.

Fudenberg, D., and J. Tirole, 1991, “Perfect Bayesian equilibrium and sequential equilibrium,” *Journal of Economic Theory*, 53(2), 236–260.

Gallien, F., S. Kassibrakis, S. Malamud, and F. Passerini, 2016, “Managing Inventory with Proportional Transaction Costs,” working paper.

Gârleanu, N., 2009, “Portfolio choice and pricing in illiquid markets,” *Journal of Economic Theory*, 144(2), 532–564.

Glode, V., and C. Opp, 2016, “Asymmetric Information and Intermediation Chains,” *American Economic Review*, 106(9), 2699–2721.

Gromb, D., and D. Vayanos, 2002, “Equilibrium and welfare in markets with financially constrained arbitrageurs,” *Journal of Financial Economics*, 66(2-3), 361–407.

——— , 2010, “Limits of Arbitrage,” *Annual Review of Financial Economics*, 2(1), 251–275.

Grossman, S. J., and M. H. Miller, 1988, “Liquidity and Market Structure,” *Journal of Finance*, 43(3), 617–37.

He, Z., and A. Krishnamurthy, 2011, “A Model of Capital and Crises,” *The Review of Economic Studies*, 79(2), 735–777.

Hendershott, T., D. Li, D. Livdan, and N. Schürhoff, 2016, “Relationship Trading in OTC Markets,” working paper, Haas, Federal Reserve System and HEC Lausanne.

Hendershott, T., and A. J. Menkveld, 2014, “Price pressures,” *Journal of Financial Economics*, 114, 405–423.

Hendershott, T., and M. S. Seasholes, 2007, "Market Maker Inventories and Stock Prices," *American Economic Review*, 97(2), 210–214.

Herskovic, B., and J. Ramos, 2015, "Information Through Peers: Coordination Impacts on Network Formation," working paper, UCLA Anderson, USC Marshall.

Hollifield, B., A. Neklyudov, and C. S. Spatt, 2014, "Bid-Ask Spreads, Trading Networks and the Pricing of Securitizations: 144a vs. Registered Securitizations," working paper, Carnegie Mellon University and HEC Lausanne.

Huang, Y., and W. F. McColl, 1997, "Analytical inversion of general tridiagonal matrices," *Journal of Physics A: Mathematical and General*, 30(22), 7919–7933.

Hugonnier, J., B. Lester, and P.-O. Weill, 2016, "Heterogeneity in decentralized asset markets," working paper, EPFL, Federal Reserve Bank of Philadelphia and UCLA.

in 't Veld, D., and I. van Lelyveld, 2014, "Finding the core: Network structure in interbank markets," *Journal of Banking & Finance*, 49, 27–40.

in 't Veld, D. L., M. van der Leij, and C. H. Hommes, 2014, "The Formation of a Core Periphery Structure in Heterogeneous Financial Networks," working paper, University of Amsterdam.

James, J., I. W. Marsh, and L. Sarno (eds.), 2012, *Handbook of Exchange Rates*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Karmel, R. S., 2008, "Should Securities Industry Self-Regulatory Organizations be Considered Government Agencies," *Stanford Journal of Law, Business & Finance*, 14.

Kirilenko, A. A., A. S. Kyle, M. Samadi, and T. Tuzun, 2011, "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market," working paper, Imperial College Business School, University of Maryland, Southern Methodist University and Federal Reserve Board.

Lagos, R., G. Rocheteau, and P.-O. Weill, 2008, "Crashes and Recoveries in Illiquid Markets," working paper, New York University, U.C. Irvine and U.C. Los Angeles.

Levin, D. A., Y. Peres, and E. L. Wilmer, 2009, *Markov Chains and Mixing Times*. American Mathematical Soc.

Li, D., and N. Schürhoff, 2014, "Dealer Networks," working paper, Federal Reserve Board and Université de Lausanne.

Mailath, G. J., A. Postlewaite, and L. Samuelson, 2005, "Contemporaneous perfect epsilon-equilibria," *Games and Economic Behavior*, 53(1), 126–140.

Markose, S., S. Giansante, and A. R. Shaghaghi, 2012, "'Too interconnected to fail' financial network of US CDS market: Topological fragility and systemic risk," *Journal of Economic Behavior & Organization*, 83(3), 627–646.

Markose, S. M., 2012, "Systemic Risk from Global Financial Derivatives: A Network Analysis of Contagion and Its Mitigation with Super-Spreader Tax," *IMF Working Papers*, 12(282).

McPartland, K., 2014, "The SEF Landscape: Beyond the Numbers," *Greenwich Report*.

Meli, J. A., 2002, "Do capital constraints on market makers matter? : Evidence from the United States treasury market," working paper, University of Hong Kong.

Mengle, D., 2010, "Concentration of OTC Derivatives among Major Dealers," *ISDA Research Notes*, (4), 1–7.

Neklyudov, A., and B. Sambalaibat, 2015, "Endogenous Specialization and Dealer Networks," working paper, HEC Lausanne and Kelley School of Business.

Neklyudov, A. V., 2014, "Bid-Ask Spreads and the Over-the-Counter Interdealer Markets: Core and Peripheral Dealers," working paper, Université de Lausanne.

Peltonen, T. A., M. Scheicher, and G. Vuillemey, 2014, "The network structure of the CDS market and its determinants," *Journal of Financial Stability*, 13, 118–133.

Protter, P. E., 2005, *Stochastic Integration and Differential Equations*, vol. 21 of *Stochastic Modelling and Applied Probability.* Springer Berlin Heidelberg, Berlin, Heidelberg.

Rinne, K., and M. Suominen, 2010, "Short Term Reversals, Returns to Liquidity Provision and the Costs of Immediacy," working paper, Aalto University School of Business and Luxembourg School of Finance.

———, 2011, "A Structural Model of Short-Term Reversals," working paper, Aalto University School of Business and Luxembourg School of Finance.

Rostek, M., and M. Weretka, 2012, "Price Inference in Small Markets," *Econometrica*, 80(2), 687–711.

Shen, J., B. Wei, and H. Yan, 2015, "Financial Intermediation Chains in a Search Market," working paper, London School of Economics, Fed Atlanta and DePaul University.

Üslu, S., 2016, "Pricing and Liquidity in Decentralized Asset Markets," working paper, Carey Business School, Johns Hopkins University.

Vives, X., 2011, "Strategic Supply Function Competition With Private Information," *Econometrica*, 79(6), 1919–1966.

Watson, J., 2016, "Perfect Bayesian Equilibrium: General Definitions and Illustrations," working paper, University of California, San Diego.

Weill, P.-O., 2007, "Leaning Against the Wind," *Review of Economic Studies*, 74(4), 1329–1354.

Zhang, F., 2010, "High-Frequency Trading, Stock Volatility, and Price Discovery," working paper, Yale School of Management.

Zhong, Z., 2014, "The Network of Inter-Dealer Trading in an Over-the-Counter Market," working paper, Department of Finance, University of Melbourne.