Shengxing Zhang

Liquidity misallocation in an over-the-counter market

Article (Accepted version)
(Refereed)

Original citation:
Zhang, Shengxing (2018) *Liquidity misallocation in an over-the-counter market.* Journal of Economic Theory, 174. pp. 16-56. ISSN 0022-0531

DOI: 10.1016/j.jet.2017.11.008

Liquidity Misallocation in an Over-The-Counter Market

Shengxing Zhang*

London School of Economics

January 17, 2018

Abstract

To understand the illiquidity of the over-the-counter market when dealers and traders are in long-term relationships, I develop a framework to study the endogenous liquidity distortions resulting from the profit-maximizing, screening behavior of dealers. The dealer offers the trading mechanism contingent on the aggregate history of his customers summarized by the asset allocation. The equilibrium distortion is type dependent: trade with small surplus breaks down; trade with intermediate surplus may be delayed; trade with large surplus is carried out with a large bid/ask spread but without delay. Because of dealers' limited commitment, the distortions become more severe when the valuation shock is frequent, the valuation dispersion is large or the matching friction to form new relationships is large. Calibrating the model and running a horse race between matching efficiency, trading speed and relationship stability, I found that the liquidity disruption in the market during the recent financial crisis is more consistent with declining matching efficiency of forming trading relationships. The optimal mechanism can be implemented by random quote posting.

Key words: *Screening; Liquidity; Long-term relationship; Over-the-counter markets*

JEL: D82, D83, G1

*Email: S.Zhang31@lse.ac.uk. I thank the support from the Centre for Macroeconomics at LSE.

1 Introduction[1]

Many financial over-the-counter (OTC) markets are illiquid, with large bid-ask spreads and long delay in trade. For example, a typical municipal bond is traded in the OTC market once every 25 days with an average bid-ask spread of more than 50 basis point[16], whereas an equity is typically traded more than once every second with bid-ask spreads an order of magnitude smaller[32]. Standard theories of the OTC market such as Duffie, Garleanu and Pedersen (2005)[9], Lagos and Rocheteau (2010)[25] and Hugonnier, Lester and Weill (2014)[22] show that liquidity distortions can arise from the search friction to locate trading counterparties. However, financial institutions typically maintain long-term relationships with each other.[2] The friction to locate trading counterparties may not be so large as to explain all the distortions. In this paper, I show that even when broker-dealers maintain long-term relationships with traders, additional liquidity distortions may arise if they do not observe traders' private valuations. In this environment, dealers screen traders by controlling the speed of trade and the transaction price and keep track of the endogenous asset allocation across his customers. Although dealers provide immediacy to traders, the monopsony power of the dealer in the OTC market leads to imperfect allocation of liquidity.

In the model, traders search and match with dealers. Similar to equilibrium search models of the labor market[33], the match is long-term but subject to breakup shocks and forming new matches takes time. Each dealer is matched with a continuum of traders. By posting menus of contracts specifying trading probabilities and transaction prices for

[1] I am indebted to Douglas Gale, Boyan Jovanovic, Ricardo Lagos and Thomas Sargent for their encouragement and comments. The paper has also benefited from discussion with Gaetano Antinolfi, Saki Bigio, Anton Braun, Kenneth Burdett, Jonathan Chiu, Benjamin Lester, Guido Menzio, William Roberds, Guillaume Rocheteau, Cecilia Siritto, Klaus-Peter Hellwig, Ariel Zetlin-Jones, Pierre-Olivier Weill, and Randall Wright. I also thank the seminar participants at the 2012 Midwest Macroeconomics Conference, the 2012 Scociety of Economic Dynamic Annual Meeting, the 2012 North American Summer Meeting of the Econometric Society, the 2012 Chicago Federal Workshop on Money, Banking, Payments, and Finance, and the Search and Matching Workshop at UPenn.

[2] See, for example, Li and Schürhoff (2014)[29] and Afonso Kovner and Schoar (2014)[1].

indivisible assets, dealers screen traders with heterogeneous gain from trade. In equilibrium, traders with large gain from trade value most immediacy. So, they are willing to pay a high premium relative to the market price to trade faster. Traders with intermediate gain from trade sacrifice trading speed for a lower spread. To induce traders with large gain to accept the high premium, dealers strategically exclude traders with small gain by charging a fixed spread on top of a variable spread increasing in the trading probability. Therefore, trade breaks down for those with small gains from trade. I show that these three types of distortions could coexist in equilibrium.

A theoretical contribution of the paper is to formalize and solve the dynamic programing problem of the dealer, in which the asset allocation to customers is a high-dimensional state variable and the contract menus are the control variable. The asset allocation is slow moving because of the strategic delay of the dealer and the physical limit on the trading frequency between the dealer and his customers. This induces the dynamic interaction between the trading mechanism and the asset allocation, especially for traders with intermediate gain from trade. Traders with large gains from trade are willing to pay a premium to trade immediately. But trading faster with them also means that less of those traders remain waiting. With less traders with large gain remaining, dealers have an incentive to trade faster with traders with intermediate gain. Meanwhile, the dealer also has limited commitment to contract menus he chooses: when he chooses date-t contract menus, he takes as given traders' reservation value, even though it depends on contract menus offered by the dealer in the future. A better deal in the future increases the reservation value, which in turn squeezes the dealer's current profit margin. The limited commitment induces dynamic competition within a long-term relationship, in the spirit of the Coase conjecture, that a dealer competes with his future self when trading with investors and this drives his monopsony rent to zero. In this rich dynamic environment, however, we will show the Coase conjecture only holds partially. Frictions in the dynamic environment, such as the persistence of the trading relationship, matching frictions to form long-term relationships, and trading speed limit, all affect market liquidity. As a result, qualitatively, for traders with intermediate gain from trade, trade may not break down but instead may be delayed. For traders with small gain, trade still breaks down as in a static screening problem.

The endogenous liquidity distortion also depends on the stability of trader-dealer relationships and the matching friction in relationship formation. In benchmark models of the OTC market such as Duffie, Garleanu and Pedersen (2005)[9] and Lagos and Rocheteau (2009)[25], liquidity distortion arises because of search frictions to locate trading counterparties. I show that when it is hard to form new matches, dealers may also be more likely to delay trade. Therefore, the strategic delay of the dealer could amplify the matching friction. On the other hand, the stability of the trader-dealer relationship, measured by the breakup rate of the relationship, may or may not induce more delay by the dealer, depending on the matching friction. When I calibrate the model to the corporate bonds market and run a horse race between relationship stability, matching efficiency and trading speed limit, a friction similar to those in benchmark models, I found that increasing the matching friction is more consistent with stylized facts from the recent financial crisis, during which dealers charge higher markups to their clients without increasing liquidity provision.[3]

The theory also links the opacity of the OTC market to market liquidity, transaction costs and asset prices. Opacity leads to heterogeneous valuation over an asset. I show that more dispersed valuations lead to more delay in trade and larger spread for all traders.

The optimal trading mechanism can be implemented by lotteries over transaction prices. So, the assumption that the dealer offers the contract menus is not restrictive, even when the dealer can only use outright sales contract. My theory gives an alternative explanation for the price dispersion in opaque markets described in Green (2007)[15].

Literature Review Since assets are durable and dealers have limited commitment to the trading mechanism they post. This paper is closely related to the literature on the Coase conjecture (see, for example, Gul, Sonnenschein and Wilson (1986)[19]). Recent developments in this literature include Fuchs and Skrzypacz (2010)[12] and Garrett (2013)[13]. My

[3] See Di Maggio, Kermani and Song (2016)[8].

model allows changing valuations and arrival of new buyers over time, as well as the frictions in forming and maintaining long-term relationships, none of which are trivial additions to the baseline setup. The literature studies the optimal trading mechanism contingent on the individual history. By summarizing by the asset allocation the aggregate history of the dealer' customers, which includes trading activities and other events in the past, the trading mechanism in the paper becomes more tractable and captures the effect of not only trading speed limit but also other frictions on the liquidity distortion in the market.

My paper is also related to the literation on strategic delay induced by asymmetric information. Guerrieri, Shimer and Wright (2010)[18] and Chang (2011)[6] combine directed search friction and asymmetric information. In these papers, delay in trade serves as a signaling device. One common issue with the signaling equilibrium is that it is sensitive to perturbation. For an environment with nearly complete information, the equilibrium allocation is the same as an environment with severe adverse selection as long as the support of the distribution remains the same. In this model, delay in trade serves as a screening device. Dealers take into account the effect of the type distribution on their profit when they optimize the menu of contracts they offer. A perturbation to the distribution will not affect qualitatively the equilibrium allocation. In this sense, the equilibrium allocation is more robust in this model. Chiu and Koeppl (2011)[7], Camargo and Lester (2011)[5], Guerrieri and Shimer (2011)[17] and Chang (2011)[6] study the effect of lemons problem on liquidity. These papers take the sales contract as given, while here I allow contracts consisting of two ingredients, the trading probability and the transaction price for sales.

Another related literature is about asymmetric information in the decentralized dynamic market. Lauermann and Wolinsky (2011)[27], Wolinsky (1990)[34] and Blouin and Serrano (2001)[3] study information aggregation in the lemon market and its social welfare as the friction vanishes. Golosov, Lorenzoni, and Tsyvinski (2009)[14] and Camargo and Lester (2011)[5] are concerned with the trading dynamics. Horner and Vieille (2009)[21] studies the interaction of strategies of sequentially arriving short-lived buyers in a dynamic lemon market. Inderst (2005)[23] studies limiting property of a matching market with adverse selection. Hendel and Lizzeri (1999)[20] studies the durable goods market with lemons. These papers study the implications of asymmetric information about the quality of the goods or assets being traded, while this paper studies the implications of asymmetric information about private valuations. Most of these papers focus on the limiting results as frictions are asymptotically zero. My focus is to characterize the equilibrium with frictions to study the resulting liquidity misallocation. In Duffie, Malamud and Manso (2014)[10], the authors characterize analytically the dynamics of information aggregation in segmented markets. I do not allow information propagation in the model. So valuations across traders do not interact. While my focus is on the liquidity distortion due to the interaction between asymmetric information and other frictions and shocks in the economy, it would be interesting to further study the interaction between these two channels: information aggregation and the liquidity distortion that arises from the dealer's screening strategies.

The rest of the paper is organized as follows. In Section 2, I lay out the model. In Section 3, I characterize the equilibrium. In Sections 4 and 5, I focus on distinct implications of the model on market liquidity and the quoting strategy of the dealer.

2 Model

Environment, Endowment and Preferences The economy is set in continuous time and lasts forever. There is a continuum of long-lived traders and dealers. The measures of both groups of agents are normalized to one. Traders are endowed with one type of asset and deep pockets of numeraire goods. An asset bears a unit flow of dividend goods. The total asset supply is $A = 0.5$. The asset is homogeneous in quality. I assume that traders can hold either zero or one unit of the asset.

Traders' valuation over the dividend good depends on their preference type x, which is her private information.[4] A trader with preference type x enjoys flow utility x from holding one

[4] The asset holding can also be thought of as unobservable. Since holding more than one unit of asset or selling more than the trader's holding does not generate any profit, it is better off for traders to reveal truthfully their asset holding.

unit of the asset. Over time, a trader may experience a preference shock, which arrives with Poisson rate δ. After the shock, she draws randomly a new preference type from a continuous distribution $G(\cdot)$ on a compact interval $[x_L, x_H] \subseteq \mathbb{R}_+$. Assume that the distribution is symmetric. Let $x_M = \frac{x_L + x_H}{2}$. Symmetric distribution means $G(x) = 1 - G(2x_M - x)$ for $x \in [x_L, x_H]$. Both the arrival time of a preference shock and the realization of the shock are idiosyncratic. Dealers do not value dividend goods. All agents are risk neutral and derive one util from consuming a numeraire good. Traders and dealers share a common discount rate, $r \in \mathbb{R}_+$.

The Structure of the OTC Market Agents trade assets for numeraire goods through an over-the-counter market. The OTC market has two tiers as in Lagos and Rocheteau (2009)[25]. Only dealers can trade in the first tier of the market, which we call the inter-dealer market. The inter-dealer market is competitive, where dealers buy and sell any amount of assets at the market price without delay. Denote the market clearing price at t to be P_t. The second tier is a bilateral market between dealers and traders. Dealers and traders match randomly. A trader can be matched with only one dealer. But the dealer is not constrained in his capacity to host customers. An unmatched trader meets a dealer with Poisson rate α. The relationship between a matched trader and her dealer breaks up at Poisson rate γ. Therefore, the trader-dealer relationship is long-term. This market structure can be thought as the stable core-periphery structure observed in OTC markets.[5]

Trade between a dealer and his customers is bilateral and takes place repeatedly in each dealer-customer relationship at Poisson rate q. The arrival rate represents the speed limit at which the dealer can process a trade, which includes the time to bargain and deliver the asset. Also, the dealer may be busy handling trades from his other customers, so each trader has to wait for her turn. Only when the trader finds the occasion that the dealer is not busy, he gets to negotiate the terms of the bilateral trade. This is also part of the physical delay represented by the Poisson rate.

[5] See, for example, Li and Schürhoff (2013)[29].

On each trading occasion, the dealer offers two feasible menus of contracts, a menu of bid contracts for traders who want to sell and a menu of ask contracts for traders who want to buy. The trader chooses one contract from the menus.

Trading Mechanism Posted by Dealers Each contract in menu i has two components, the probability of trade, denoted by $q_{imt}/q \in [0,1]$, and the transaction price, denoted by $p_{imt} \in \mathbb{R}_+$, where $i \in \{a, b\}$ indicates whether it is an asking contract or a biding contract and $m \in \mathbb{X} \equiv [x_L, x_H]$ indicates the reported preference type of the corresponding customer. A contract menu $i \in \{a, b\}$, denoted by $\mathbb{M}_{it} = \{(q_{imt}, p_{imt})\}_{\forall m \in \mathbb{X}}$, is *feasible* if (1) given the menu, it is incentive compatible (IC) for traders to report truthfully their preference types; (2) it is individual rational (IR) for a trader of type x to accept ex ante a contract with index x.

The trading mechanism offered by the dealer is contingent on the *aggregate history* for the group of customers matched with him, summarized by the asset allocation to his trader customers. The summary statistics for the group of traders is analog to the posterior belief about the preference type of a trader as the summary statistics for the individual history, if the mechanism were contingent on the individual history. If the dealer trades faster with traders of a certain type before t, there will be less trading request from traders of that type at t. The summary statistics also includes other historical information, such as the arrival of new traders, the breakup of existing relationships and the arrival of preference shocks.

If the dealer has perfect information about the history of a customer, the contract menu should be further contingent on the history. In the model, I focus on the aggregate history for the following reasons. (1) While dealers can observe the history of a trading account, a customer can hide his history by opening several accounts with the same dealer at negligible cost. She can also hide the timing at which she experiences a preference shock. These hidden actions prevent the dealer from observing the history of his customers. (2) Trading mechanisms contingent on the history of each customer, which includes not only the trading history but also the history of preference changes, the date of starting the long-term relationship and the prior of her preference type, may require too much sophistication, especially because a dealer hosts a continuum of customers.

The contract menus can be viewed as a direct mechanism, which can have alternative implementations where traders do not have to report explicitly their types. One implementation is to use price lotteries on each trading occasion, which fits well the institutional detail of brokers or market makers intermediating trade in the OTC market. I will show in Section 5 that by posting quotes randomly according to optimally chosen distributions, dealers can still the direct trading mechanism.

3 Equilibrium Definition and Existence

3.1 Equilibrium Definition

In this section, I will first present the problems of the trader and the dealer, followed by other equilibrium conditions. The equilibrium definition follows in the end. I will focus on the symmetric equilibrium.

Traders' Problem Denote the maximum attainable utility of a trader of preference type x with a units of asset and matched with a dealer at t to be M_{axt}, where $a \in \{0, 1\}$. At any moment, the trader may face several contingencies: she may receive a trading opportunity from the dealer, a preference shock or a breakup shock.

When the trader receives a trading opportunity, she picks a contract from the menu offered by the dealer. Denote the reservation value of a trader matched with a dealer by $d_{xt} \equiv M_{1xt} - M_{0xt}$. For a trader with an asset, she chooses a contract from bid contract menu, $\mathbb{M}_{bt}$.[6] The feasibility of the contract menu requires that for trader of type $1x$,

$$
\begin{aligned}
x &\in \arg\max_{m\in[x_L,x_H]} \frac{q_{bmt}}{q}\left(p_{bmt}-d_{xt}\right), && (1)\\
0 &\leq \frac{q_{bxt}}{q}\left(p_{bxt}-d_{xt}\right), && (2)
\end{aligned}
$$

where equation (1) is the incentive compatibility (IC) constraint for the trader, equation (2) is the ex ante individual rationality (IR) constraint. The expected value for trader of type $1x$ to pick contract (q_{bxt}, p_{bxt}) at t is $\frac{q_{bxt}}{q}\left(M_{0xt}+p_{bxt}\right)+\left(1-\frac{q_{bxt}}{q}\right)M_{1xt}$.

If she receives a preference shock, she draws her new preference type from distribution $G(x)$. So her expected utility in this case is $\int M_{1\tilde{x}t}dG(\tilde{x})$.

If she face an exogenous breakup shock, her expected utility is the maximum attainable utility of a trader of type $1x$ not matched with a dealer, which is denoted $U_{1x\tau}$. Although a trader has the option to break up at any moment with the dealer matched with her, she never has the incentive to do so because any feasible contract menu offered by a dealer satisfies her participation constraint and therefore makes her weakly better off than breaking up with him. We will see that in a symmetric equilibrium the set of feasible contract menus is nonempty.

Given the maximum attainable utility on these contingencies, we can apply Bellman's principle of optimality and write M_{1xt} as

$$
\begin{aligned}
M_{1xt} = \mathbb{E}_t\Bigg\{\int_t^\tau e^{-r(s-t)}x ds + e^{-r(\tau-t)}\Bigg[&\mathbb{I}_{\{\tau=\tau_\delta\}}\int M_{1\tilde{x}\tau}dG(\tilde{x}) && (3)\\
&+\mathbb{I}_{\{\tau=\tau_\gamma\}}U_{1x\tau}\\
&+\mathbb{I}_{\{\tau=\tau_q\}}\left(\frac{q_{bx\tau}}{q}\left(M_{0x\tau}+p_{bx\tau}\right)+\left(1-\frac{q_{bx\tau}}{q}\right)M_{1x\tau}\right)\Bigg]\Bigg\}
\end{aligned}
$$

where τ_δ is an exponential random variable with parameter δ that represents the arrival of preference shock, τ_γ is an exponential random variable with parameter γ that represents the arrival of breakup shock, τ_q is an exponential random variable with parameter q that represents the arrival of a trading opportunity and $\tau \equiv \min\{\tau_\delta, \tau_\gamma, \tau_q\}$.

[6] Because the trader can only hold zero or one unit of asset, she gains nothing from misreporting her asset holding.

The utility of type $1x$ trader can be simplified using trader's reservation value as

$$\begin{aligned} M_{1xt} \;=\; \mathbb{E}_t \Bigg\{ \int_t^{\tau} e^{-r(s-t)} x ds + e^{-r(\tau-t)} \Bigg[& M_{1x\tau} \\ & + \mathbb{I}_{\{\tau=\tau_\delta\}} \int \left(M_{1\tilde{x}\tau} - M_{1x\tau}\right) dG(\tilde{x}) \\ & + \mathbb{I}_{\{\tau=\tau_\gamma\}} \left(U_{1x\tau} - M_{1x\tau}\right) \\ & + \mathbb{I}_{\{\tau=\tau_q\}} \frac{q_{bx\tau}}{q} \left(p_{bx\tau} - d_{x\tau}\right) \Bigg] \Bigg\} \end{aligned} \tag{4}$$

Similarly, feasibility of the contract menu, $\mathbb{M}_{at}$, requires that for trader of type $0x$,

$$x \;\in\; \arg \max_{m \in [x_L, x_H]} \frac{q_{amt}}{q} \left(d_{xt} - p_{amt}\right), \tag{5}$$

$$0 \;\leq\; \frac{q_{axt}}{q} \left(d_{xt} - p_{axt}\right), \tag{6}$$

where equation (5) is the incentive compatibility (IC) constraint for the trader, equation (6) is the ex ante individual rationality (IR) constraint. And the maximum attainable utility of a trader who does not own an asset and is matched with a dealer satisfies

$$\begin{aligned} M_{0xt} \;=\; \mathbb{E}_t \Bigg\{ e^{-r(\tau-t)} \Bigg[& M_{0x\tau} + \mathbb{I}_{\{\tau=\tau_\delta\}} \int \left(M_{0\tilde{x}\tau} - M_{0x\tau}\right) dG(\tilde{x}) \\ & + \mathbb{I}_{\{\tau=\tau_\gamma\}} \left(U_{0x\tau} - M_{0x\tau}\right) \\ & + \mathbb{I}_{\{\tau=\tau_q\}} \frac{q_{ax\tau}}{q} \left(d_{x\tau} - p_{ax\tau}\right) \Bigg] \Bigg\}, \end{aligned} \tag{7}$$

where $\tau \equiv \min\{\tau_\delta, \tau_\gamma, \tau_q\}$. Subtracting (7) from (4) shows that the reservation value satisfies the following equation

$$\begin{aligned} d_{xt} \;=\; \mathbb{E}_t \Bigg\{ \int_t^{\tau} e^{-r(s-t)} x ds + e^{-r(\tau-t)} \Bigg[& d_{x\tau} \\ & + \mathbb{I}_{\{\tau=\tau_\delta\}} \int \left(d_{\tilde{x}\tau} - d_{x\tau}\right) dG(\tilde{x}) \\ & + \mathbb{I}_{\{\tau=\tau_\gamma\}} \left(w_{x\tau} - d_{x\tau}\right) \\ & + \mathbb{I}_{\{\tau=\tau_q\}} \frac{1}{q} \left(q_{ax\tau} p_{ax\tau} + q_{bx\tau} p_{bx\tau} - \left(q_{ax\tau} + q_{bx\tau}\right) d_{x\tau}\right) \Bigg] \Bigg\}, \end{aligned} \tag{8}$$

where $w_{xt} \equiv U_{1xt} - U_{0xt}$ denotes the reservation value of a trader not matched with a dealer, U_{axt} denotes the utility of a trader who owns a unit of the asset and is not matched with a dealer.

$$U_{1xt} = \mathbb{E}_t \left\{ \int_t^{\tau} e^{-r(s-t)} x ds + e^{-r(\tau-t)} \left[\mathbb{I}_{\{\tau=\tau_\delta\}} \int U_{1\tilde{x}\tau} dG(\tilde{x}) + \mathbb{I}_{\{\tau=\tau_\alpha\}} M^*_{1x\tau} \right] \right\}, \tag{9}$$

where τ_δ is an exponential random variable with parameter δ that represents the arrival of preference shock, τ_α is an exponential random variable with parameter α that represents the arrival of a dealer, $M^*_{1x\tau}$ denotes the equilibrium expected utility from being matched with a dealer, and $\tau \equiv \min\{\tau_\delta, \tau_\alpha\}$. For the rest of the paper, I use the "$*$" superscript to represent equilibrium values that an agent takes as given and does not immediately control, wherever it clarifies presentation.

Likewise, the utility of a trader who owns an asset and is not matched with a dealer, denoted U_{0xt}, satisfies

$$U_{0xt} = \mathbb{E}_t \left\{ e^{-r(\tau-t)} \left[\mathbb{I}_{\{\tau=\tau_\delta\}} \int U_{0\tilde{x}\tau} dG(\tilde{x}) + \mathbb{I}_{\{\tau=\tau_\alpha\}} M^*_{0x\tau} \right] \right\}. \tag{10}$$

where $\tau \equiv \min\{\tau_\delta, \tau_\alpha\}$.

In a symmetric equilibrium where dealers employ the same strategy at any moment, $M^*_{1x\tau} = M_{1x\tau}$, $M^*_{0x\tau} = M_{0x\tau}$, on the equilibrium path. Subtracting (10) from (9), we derive an equation for the reservation value of a trader not matched with a dealer.

$$w_{xt} = \mathbb{E}_t \left\{ \int_t^{\tau} e^{-r(s-t)} x ds + e^{-r(\tau-t)} \left[\mathbb{I}_{\{\tau=\tau_\delta\}} \int w_{\tilde{x}\tau} dG(\tilde{x}) + \mathbb{I}_{\{\tau=\tau_\alpha\}} d_{x\tau} \right] \right\}. \tag{11}$$

To restrict attention to solutions to the above problem relevant to our economic analysis, we further require that the utility of traders and their reservation value solved by equations are uniformly bounded.[7] This implies the utility functions and reservation value satisfy

[7] If a trader of type x holds on to her asset forever and she never experiences preference shocks, her life time utility would be x/r. A trader cannot receive a lower payoff than x/r through trading and receiving upward preference shocks if her current preference type is equal to x_L and she cannot receive a higher payoff than x/r through trading and receiving downward preference shocks if her current preference type is equal to x_H. Therefore, we would expect $x_H/r \geq M_{1xt}, U_{1xt} \geq x_L/r$. As long as the asset pays nonnegative dividend, we would also expect that $x_H/r \geq M_{0xt} U_{0xt} \geq 0$.

transversality conditions,

$$\lim_{t\to\infty} e^{-rt} M_{axt} = \lim_{t\to\infty} e^{-rt} U_{axt} = \lim_{t\to\infty} e^{-rt} d_{xt} = \lim_{t\to\infty} e^{-rt} w_{xt} = 0, \forall (a,x) \in \{0,1\} \times \mathbb{X}. \quad (12)$$

To further characterize the payoff functions, it is necessary to study some properties of contract menus offered by the dealer.

A feasible contract menu satisfies the following properties.

Lemma 1. *Given any nonnegative reservation value function d_{xt}, a feasible contract menu at time t corresponding to the reservation value function satisfies the following properties: For any two reservation value d_{xt}, d_{yt} such that $d_{xt} > d_{yt}$. The corresponding contracts picked by type x and y has the following properties:*

1. $q_{axt} \geq q_{ayt}$, $q_{bxt} \leq q_{byt}$;
2. $q_{axt}p_{axt} \geq q_{ayt}p_{ayt}$, $q_{bxt}p_{bxt} \leq q_{byt}p_{byt}$;
3. $q_{axt}(d_{xt} - p_{axt}) \geq q_{ayt}(d_{yt} - p_{ayt})$, $q_{bxt}(d_{xt} - p_{bxt}) \geq q_{byt}(d_{yt} - p_{byt})$.
4. *If $q_{axt}q_{bxt} = 0$, for all x, there exists a cutoff κ_t such that is $q_{axt} = 0$ if $x < \kappa_t$ and $q_{bxt} = 0$ if $x > \kappa_t$.*

With Lemma 1, we can prove the following properties of traders' payoff functions and reservation value functions.

Lemma 2. *Assume $q_{axt}q_{bxt} = 0$ for all x. Then, under boundedness conditions (12), given symmetric and feasible contract menus, there exist unique uniformly bounded functions $d : [x_L, x_H] \times \mathbb{R}_+ \to \mathbb{R}$ and $w : [x_L, x_H] \times \mathbb{R}_+ \to \mathbb{R}$ that satisfy (8) and (11). They are absolutely continuous in $(x,t) \in [x_L, x_H] \times \mathbb{R}_+$ and strictly increasing in $x \in [x_L, x_H]$ with a uniformly bounded derivative with respect to type. Given d_{xt} and w_{xt}, there exist unique functions M_{1xt}, M_{0xt}, U_{1xt} and U_{0xt} that are uniformly bounded and satisfy (4), (7), (9), (10).*

I impose that $q_{axt}q_{bxt} = 0$ for all x because as we will see in the dealer's problem, it is a constraint the dealer faces. It means that the dealer does not engage in round-trip trading, to buy and sell simultaneously with traders of the same preference type.

My method for solving for values and establishing their properties is borrowed extensively from Hugonnier, Lester and Weill (2014) [22]. In Hugonnier, Lester and Weill (2014) [22], division of trade surplus and trading frequency between two types are exogenous because of symmetric information. Here, asymmetric information between the dealer and the trader makes the division of trade surplus and trading frequencies endogenous. My proof handled additional technicalities induced by these features.

Laws of Motion Denote the population density functions of traders matched with a dealer to be n^m_{axt} and the density functions of those not matched to be n^u_{axt}, where $a \in \{0, 1\}$. In a symmetric equilibrium, the laws of motion of these density functions are the following

$$\dot{n}^m_{1xt} = \alpha n^u_{1xt} - \delta n^m_{1xt} + \delta g(x) N^m_{1t} - \gamma n^m_{1xt} - q_{bxt} n^m_{1xt} + q_{axt} n^m_{0xt}, \tag{13}$$

$$\dot{n}^u_{1xt} = -\alpha n^u_{1xt} - \delta n^u_{1xt} + \delta g(x) N^u_{1t} + \gamma n^m_{1xt}, \tag{14}$$

$$\dot{n}^m_{0xt} = \alpha n^u_{0xt} - \delta n^m_{0xt} + \delta g(x) N^m_{0t} - \gamma n^m_{0xt} + q_{bxt} n^m_{1xt} - q_{axt} n^m_{0xt}, \tag{15}$$

$$\dot{n}^u_{0xt} = -\alpha n^u_{0xt} - \delta n^u_{0xt} + \delta g(x) N^u_{0t} + \gamma n^m_{0xt}, \tag{16}$$

where $N^k_{At} \equiv \int n^k_{Axt} dx$ denotes the total measure of traders with A unit of the asset and matching status $k \in \{m, u\}$. Apart from the inflows and outflows induced by exogenous events such as matching or breaking up with a dealer, and receiving a preference shock, the trading frequencies picked by dealers, q_{bxt} and q_{axt}, also affect the asset allocation for traders matched with a dealer. The delay in trade specified in dealers' contract menus also imposes an externality on the asset allocation to unmatched traders. This affects the strategic interaction across dealers because dealers take the inflow from unmatched traders as given.[8]

Dealers' Problem A dealer's maximum attainable payoff depends on the expected profit from each contract and the distribution of contract choices by his customers. Dealer's prob-

[8]Notice also that the law of motion of $n^m_{1xt} + n^m_{0xt}$ does not depend on dealers' trading strategy. This will be explained in more detail in the dealer's problem.

lem at period t is the following:

$$D_{t,t} \equiv \max_{\substack{\{\mathbb{M}_{as}, \mathbb{M}_{bs}\}_{s\geq t} \\ \{n^m_{1\cdot s}\}_{s>t}}} \int_t^\infty e^{-r(s-t)}[q_{axs}(p_{axs} - P_s)n^m_{0xs} + q_{bxs}(P_s - p_{bxs})n^m_{1xs}]dxds, \tag{17}$$

subject to, for all $x \in [x_L, x_H]$ and $s \geq t$,

$$\dot{n}^m_{1xs} = \alpha n^{u*}_{1xs} - \delta n^m_{1xs} + \delta g(x) N^m_{1s} - \gamma n^m_{1xs} - q_{bxs} n^m_{1xs} + q_{axs}(n^m_{xs} - n^m_{1xs}), \tag{18}$$

$$\dot{n}^m_{xs} = \alpha n^{u*}_{xs} - \delta n^m_{xs} + \delta g(x) N^m_s - \gamma n^m_{xs}, \tag{19}$$

$$q_{axs} q_{bxs} = 0, \tag{20}$$

$\mathbb{M}_{as}$,$\mathbb{M}_{bs}$ are feasible.

In the objective function of the profit maximizing dealer, what differentiates dynamic screening from static screening is that the choice of the screening contract menu affects the law of motion of the state variable, n^m_{1xt}. The density function, n^m_{1xt}, summarizes the trading history between the dealer and his customer-traders. Its law of motion is characterized by (18). If he sells assets to traders of type x at a higher frequency, more assets will be allocated to type x traders. If the dealer buys assets at a higher frequency, less assets will be allocated to them. The asset allocation in turn affects the equilibrium demand and supply of assets from traders and therefore the dealer's incentive to screen traders. The density function also contains other historical information, such as the arrival of new traders, the breakup of existing relationships and the arrival of preference shocks. n^{u*}_{Axt}, for $A \in \{0, 1\}$, denotes the equilibrium type distributions of traders with A unit of the asset and unmatched with the dealer. When the dealer chooses his contract menus, he takes as given the type distributions of inflows from the pool of unmatched traders, αn^{u*}_{Axs}. On the equilibrium path of the symmetric equilibrium, n^{u*}_{0xt} and n^{u*}_{1xt} follow laws of motion (14) and (16). $N^m_{At} = \int n^m_{Axt} dx$, for $A \in \{0, 1\}$. I implicitly impose in the dealer's problem that he does not willingly breakup with the traders matched with him, even when he considers his deviation strategies. This is because he does not face capacity constraints in terms of hosting customers. So, the opportunity cost of retaining a customer is zero. Then, his payoff from retaining a customer always weakly dominates that from breaking up with her. Since the dealer does not break up with his customers, the population density of his customers unconditional on asset allocation, $n^m_{xt} \equiv n^m_{1xt} + n^m_{0xt}$, follows an exogenous law of motion, (19), which depends only on the inflow of customers, exogenous breakups and preference shocks. $n^{u*}_{xt} \equiv n^{u*}_{1xt} + n^{u*}_{0xt}$, and $N^m_t \equiv \int n^m_{xt} dx$. Because asset quality is homogeneous, round-trip trading with traders of the same preference type cannot be at the same time profitable for the dealer and individual rational for those traders. I will conjecture and verify that constraint (20) is not binding in equilibrium.

Denote $\kappa_t \in [x_L, x_H]$ to be the cutoff type so that $d^*_{\kappa_t t} = P_t$. κ_t exists because otherwise, either demand or supply in the interdealer market will be zero, so that the interdealer market cannot clear. There exists a unique κ_t corresponding to the reservation value function d_{xt} because d_{xt} is strictly increasing in x, according to Lemma 2.

Lemma 3. *Taking as given the cutoff type κ_t, a sufficient and necessary condition for the contract menus to be incentive compatible is that trading frequencies for buyers, q_{axt}, are increasing in the preference type and trading frequencies for sellers, q_{bxt}, are decreasing in the preference type, and prices in the equilibrium contract menu follow the following equations,*

$$q_{bxt} p_{bxt} = q_{bxt} d^*_{xt} + \int_x^{\kappa_t} q_{bst} d^{*\prime}_{st} ds, \tag{21}$$

$$q_{axt} p_{axt} = q_{axt} d^*_{xt} - \int_{\kappa_t}^x q_{ast} d^{*\prime}_{st} ds. \tag{22}$$

Lemma 3 means that the feasibility constraints on date-t contract menus can be replaced by monotonicity constraints on q_{bxt} and q_{axt} and equations (21) and (22). Substituting equations (21) and (22) to the dealer's objective function, we have

$$D_{t,t} = \max_{\substack{\{q_{axs}, q_{bxs}\}_{s \geq t, x \in [x_L, x_H]} \\ \{n^m_{1xs}\}_{s > t, x \in [x_L, x_H]}}} \int_t^\infty \mathcal{D}_{s,t} ds \tag{23}$$

where

$$\mathcal{D}_{s,t} = e^{-r(s-t)} \int \Bigg[(q_{axs} d^*_{xs} - \int_{\kappa_s}^{x} q_{aut} d^{*\prime}_{us} du - q_{axs} P^*_s)(n^m_{xs} - n^m_{1xs}) \\ +(q_{bxs} P^*_s - q_{bxs} d^*_{xs} - \int_{x}^{\kappa_s} q_{bus} d^{*\prime}_{us} du) n^m_{1xs} \Bigg] dx \tag{24}$$

subject to,

$$(18) \text{ and } (19),$$

$$q_{bxs} \text{ is decreasing in } x, \tag{25}$$

$$q_{axs} \text{ is increasing in } x. \tag{26}$$

Because the market price, reservation values and trading frequencies are all totally bounded, the following transversality condition holds, $\lim_{s\to\infty} \mathcal{D}_{s,t} = 0$. So, we can apply the maximum principle to solve for the dealer's problem. Because the dealer takes as given investor's reservation value, the objective function is linear in trading frequencies. It is clear that the dealer's problem can be reduced to picking trading frequencies only, with density functions that are implicit functions of these trading frequencies. Because for any convex combinations of two sequences of $\{q_{axs}, q_{bxs}\}_{s\geq t}$ that satisfy constraints (18), (19), (25) and (26) also satisfy these constraints, the choice set for the control variables, trading frequencies, is then convex. Given this property, we verify in Section A.4 of the Appendix that necessary conditions provided by the maximum principle are also sufficient for optimality. The sufficiency proof is derived from Liberzon (2012)[30].

To gain more intuition about the dealer's incentive, I present below the law of motion of the co-state variable, $\lambda_{xs,t}$, for the state variable, n^m_{1xs}, of the optimal control problem ignoring monotonicity constraints and boundary constraints that $0 \leq q_{axs}, q_{bxs} \leq q$[9]. $\lambda_{xs,t}$

[9]These constraints are included in the complete characterization of the dealer's problem in Section A.4 of the Appendix.

represents the dealer's present value of a customer of type $1x$ at date-s.

$$
\begin{aligned}
-\dot{\lambda}_{xs,t} = {} & e^{-r(s-t)}\left[q_{bxs}(P_s - d^*_{xs}) - \int_x^{x_H} q_{bus} d^{*\prime}_{us} du\right] \\
& - e^{-r(s-t)}\left[q_{axs}(d^*_{xs} - P_s) - \int_{x_L}^{x} q_{aus} d^{*\prime}_{us} du\right] \\
& - \gamma\lambda_{xs,t} - \delta\lambda_{xs,t} + \delta \int \lambda_{zs,t} dG(z) \\
& - (q_{bxs} + q_{axs})\lambda_{xs,t}.
\end{aligned}
\tag{27}
$$

The left-hand side of the equation represents intuitively the flow payoff from having a customer of preference type $1x$ minus the capital gain from changes in the value of the customer.[10] The first component on the right hand side is the dealer's payoff from offering the contract menu to sellers in a static setting, discounted by $e^{-r(s-t)}$. $q_{bxs}(P_s - d^*_{xs})$ is the flow surplus from trade generated by contract (q_{bxs}, p_{bxs}). $\int_x^{x_H} q_{bus} d^{*\prime}_{us} du$ is the information rent enjoyed by a seller of preference type x. Having one more type $1x$ trade means having one less type $0x$ trade, since the total population density of customers of preference type x is given at $n^m_{xs} = n^m_{1xs} + n^m_{0xs}$. The second component on the right hand side is the loss from that in a static setting. The third component, $-\gamma\lambda_{xs,t}$, is the loss from relationship breakup. The fourth component, $-\delta\lambda_{xs,t} + \delta \int \lambda_{zs,t} dG(z)$, is the loss and gain when the trader experiences a preference shock. The dealer loses a customer of preference type $1x$ but gains a type $1z$ customer, with z following distribution $G(z)$. The last component, $-(q_{bxs} + q_{axs})\lambda_{xs,t}$, is the expected opportunity cost of trading with the customer at trading frequency $q_{bxs} + q_{axs}$, rather than trading with her later.

The dealer has limited commitment to contract menus he chooses, in that when he chooses date-t contract menus, he takes as given traders' reservation value, d^*_{xt}, even though it depends on contract menus offered by the dealer in the future. A better deal in the future increases the reservation value, which in turn squeezes the current profit margin of the dealer. The limited commitment induces dynamic competition within a long-term relationship, in the spirit of the Coase conjecture. In this rich dynamic environment, however, we will show

[10] $\dot{\lambda}_{xs} = -r\lambda_{xs}$ in the steady state.

the Coase conjecture only holds partially. Frictions in the dynamic environment, such as the persistence of the trading relationship, γ, matching frictions in the market, α, and trading speed limit, q, also affect market liquidity.

Market Clearing Condition The flow demand and supply of the asset in the inter-dealer market at any moment is equal to the bid and ask transaction flows respectively. So, the market clearing condition is

$$\int q_{axt} n^m_{0xt} dx = \int q_{bxt} n^m_{1xt} dx. \tag{28}$$

Definition 1. Given the initial distribution of traders, a *symmetric equilibrium* is value functions, $\{M_{Axt}, U_{Axt}, D_{t,t}(\cdot)\}_{A\in\{0,1\}}$, laws of motion for population distribution $\{n^m_{Axt}, n^u_{Axt}\}_{A\in\{0,1\}}$, contract menus, $\mathbb{M}_{at}$ and $\mathbb{M}_{bt}$ for all $t \in \mathbb{R}_+$, and market prices $\{P_t\}_{\forall t\in\mathbb{R}_+}$ of the inter-dealer market such that for all $A \in \{0,1\}, x \in [x_L, x_H], \tau \geq t$,

1. given the value functions of unmatched traders $U_{Ax\tau}$, the value functions of matched traders $M_{Ax\tau}$ solve traders' problem;

2. given the value functions of matched traders, the value functions of unmatched traders solve equations (9) and (10);

3. given $n^u_{Ax\tau}$ and $U_{Ax\tau}$, the contract menus $\mathbb{M}_{a\tau}, \mathbb{M}_{b\tau}$ and distribution $n^m_{Ax\tau}$ solve dealers' problem at date-t;

4. given the equilibrium contract menus, $\{n^m_{Ax\tau}, n^u_{Ax\tau}\}$ satisfy the laws of motion specified in equations (15), (16), (13) and (14);

5. the inter-dealer market clears at τ, with the market clearing condition specified in equation (28).

3.2 Characterization of the Stationary Equilibrium

I will focus on the characterization of the stationary equilibrium. In a stationary equilibrium, $\dot{n}^j_{axt} = 0$, $\forall i \in \{0,1\}, j \in \{m,u\}$, $\dot{M}_{1xt} = \dot{M}_{0xt} = \dot{U}_{1xt} = \dot{U}_{0xt} = \dot{D}_{t,t} = 0, \forall x, t$ and the equilibrium contract menus are not time varying. So I omit the time index in the notation from now on.

Conjecture 1. *In the symmetric stationary equilibrium, $\kappa = x_M$, $P = d_\kappa$, $q_{ax} = q_{b,2\kappa - x}$ for all x, $q_{ax} = 0$ for all $x < \kappa$, $q_{bx} = 0$ for all $x > \kappa$.*

Because the preference distribution is symmetric and the total asset supply is one half, it is natural to conjecture that in equilibrium $\kappa = x_M$, $P = d_\kappa$ and $q_{ax} = q_{b,2\kappa-x}$. In addition, Conjecture 1 imposes that $q_{ax} = 0$ for all $x < \kappa$, $q_{bx} = 0$ for all $x > \kappa$, which means that a dealer finds it suboptimal to trade with those she would not trade with under complete information.[11] Conjecture 1 is a sufficient condition for satisfying constraint (20). It separates the screening problem into two subproblems, one for the bid contract menu, the other for the ask contract menu. It will be verified in the proof of Theorem 1.

To present the solution to the optimal control problem, we focus on ask contracts. We start by looking at the marginal gain for a dealer from increasing trading frequency with a trader of type $0x$. Denote the marginal value if he ignores the monotonicity constraint that q_{ax} is increasing in x, and the boundary constraints that $0 \leq q_{ax} \leq q$ to be $\hat{v}_x$.[12,13]

$$\hat{v}_x \equiv \frac{r}{n^m_{0x}} \frac{\partial \mathcal{L}_t}{\partial q_{ax}}, \forall x \geq \kappa,$$

where $\mathcal{L}_t$ refers to the Lagrangian of the date-t problem of the dealer, ignoring monotonicity constraints and constraints that $0 \leq q_{ax} \leq q$.

$$\mathcal{L}_t \equiv \int_t^\infty \left\{ \mathcal{D}_{s,t} + \int_\kappa^{x_H} \lambda_{ys,t} \left[\alpha n^u_{1ys} - (\delta + \gamma)\, n^m_{1ys} + \delta g(y) \int n^m_{1zs} dz + q_{ay} n^m_{0x} - \dot{n}^m_{1ys} \right] dy \right\} ds.$$

[11] Following Property 4 of Lemma 1, increasing q_{ax} would only tighten incentive constraints of $1y$ type investors with $y > x$. So, if the dealer chooses q_{ax} to be zero under complete information, she should do so under asymmetric information. So, the conjecture is $q_{ax} = 0$ for all x with $d_x < P$, and $q_{bx} = 0$ for all x with $d_x > P$.

[12] Because we conjecture that $q_{ax} = 0$ for all $x < \kappa$, we focus on type $x \geq \kappa$ here.

[13] To single out the marginal effect on the profit from trading with a certain type of trader, for a function $J_t = \int_t^\infty \int f(y,s) dy ds$, $\frac{\partial J_t}{\partial x} \equiv \int_t^\infty \int \delta(y - x) \frac{\partial}{\partial y} f(y,s) dy ds$, which involves multiplying the integrand for the integral over y by a Dirac delta function, $\delta(y - x)$.

I first use $\hat{v}_x$ to solve for optimal trading frequencies ignoring monotonicity constraints. Then, if monotonicity constraints are violated, I apply a standard procedure called convexficiation to solve for optimal trading frequencies respecting monotonicity constraints as well.

From the first order conditions of the optimal control problem in Section A.4 of the Appendix,

$$\hat{v}_x = \frac{\int_\kappa^x (r+\delta+\gamma+q_{au}) d'_u du}{r+\delta+\gamma+q_{ax}} - \frac{\int_x^{x_H} n_{0u}^m du}{n_{0x}^m} d'_x, \forall x \geq \kappa, \text{ where} \tag{29}$$

$$d'_x = \frac{r+\alpha+\delta+\gamma}{(r+\alpha+\delta)(r+\gamma+\delta+q_{ax}) - \alpha\gamma}, \text{ and}$$

$$n_{0x}^m = \frac{\alpha}{\alpha+\gamma} \frac{\alpha+\delta+\gamma}{\alpha+\delta+\gamma+\frac{\alpha+\delta}{\delta} q_{ax}} (1-A) g(x).$$

$\hat{v}_\cdot$ is a nonlinear function of trading frequency function $q_{a\cdot}$. Intuitively, the first component in (29) represents the marginal value of increasing trading frequencies with trader of type x under complete information. The second component represents the externality of increasing the trading frequency on the dealer's profit from trading with other traders. When the dealer solves for date-t trading frequencies in the dynamic mechanism design problem, he takes as given the distribution of traders' preferences, traders' value function and the costate variable. This subproblem is equivalent to a static optimal mechanism design problem. So, the procedure in Myerson(1981)[31], which studies optimal auction design in a static environment, applies.

If $\hat{v}_x$ is increasing in x for all increasing functions,

$$q_{a\cdot} \in K \equiv \{q_{a\cdot} : [\kappa, x_H] \to [0, q], q_{ax} \text{ is increasing in } x\},$$

then trading frequency function $q_{a\cdot}$ is a fixed point of correspondence $\varphi : K \rightrightarrows K$ defined as

$$\varphi(q_{a\cdot}) \equiv \{q \in K : q_x = 0, \quad \text{if } \hat{v}_x(q_{a\cdot}) < 0,$$
$$q_x \in [0, q], \text{ if } \hat{v}_x(q_{a\cdot}) = 0,$$
$$q_x = q, \quad \text{if } \hat{v}_x(q_{a\cdot}) > 0\}.$$

Then, if there exists a fixed point, it satisfies the monotonicity constraint and the boundary constraints.

But the monotonicity of $\hat{v}_x$ is not guaranteed. In this case, construction of the correspondence involves bunching, which can be solved using a standard procedure of "convexification" following Myerson(1981)[31].

Definition 2. A function $C_x : [x_L, x_H] \to \mathbb{R}$ is a *convexification* of function $\hat{C}_x : [x_L, x_H] \to \mathbb{R}$ if for any $x \in [\kappa, x_H]$,

$$C_x \equiv \min\{\omega \hat{C}_y + (1-\omega)\hat{C}_z : \omega y + (1-\omega) z = x, \forall \omega \in [0,1], y, z \in [\kappa, x_H]\}.$$

In Section A.4 of the Appendix, we show that if C_x is a convexification of $\hat{C}_x \equiv \int_{x_L}^{x} \hat{v}_y dy$, C_x is continuously differentiable. Let $v_x \equiv C'_x$. v_x is a continuous and increasing function for all $q_{a\cdot} \in K$. The correspondence taking into consideration the monotonicity constraint is

$$\begin{aligned} \varphi(q_{a\cdot}) \equiv \{q \in K : \; & q_x = 0, && \text{if } v_x(q_{a\cdot}) < 0, \\ & q_x \in [0, q], && \text{if } v_x(q_{a\cdot}) = 0, \\ & q_{ax} = q, && \text{if } v_x(q_{a\cdot}) > 0\} \end{aligned} \tag{30}$$

In Section A.5 of the Appendix, we show that there exists a fixed point of the correspondence. Since the convexification procedure takes care of the monotonicity of trading frequencies and solves for the constrained optimal trading frequencies, the fixed point in trading frequencies is optimal and respects the monotonicity constraint. Other equilibrium objects can be derived from the fixed-point trading frequencies. Hence, the following theorem holds.

Theorem 1. *If the density function of the preference shock, $g(x)$, is continuous, there exists a stationary symmetric equilibrium, where $\kappa = x_M$, and $q_{ax} = q_{b,2x_M - x}$ for all $x \in [x_L, x_H]$.*

While Theorem 1 proves the existence of a symmetric equilibrium, the equilibrium is not necessarily unique and trading frequencies in equilibrium contract menus are not necessarily continuous in the preference type. In sections A.7 and A.8 of the Appendix, I give an example where there exist two equilibria, one in which trading frequency functions are continuous, the other in which they are step functions. In the second type of equilibrium, the endogenous type distribution induces a non-monotonic marginal value $\hat{v}_x$, resulting in bunching. For the rest of the paper, however, I will only analyze the equilibrium where trading frequencies are continuous in the preference type. For numerical and analytical examples I use, such an equilibrium exists.

4 Implications on Market Liquidity

In this section, I will first present qualitative implications of the model on market liquidity. Then, I will show the effect of dispersion of valuation, frictions in forming and maintaining long-term trader-dealer relationships on market liquidity.

4.1 Qualitative Implications

Proposition 1. *There exists an equilibrium with threshold types $\underline{x}^b$, $\bar{x}^b$, κ, $\underline{x}^a$ and $\bar{x}^a$, such that $x_L < \underline{x}^b \leq \bar{x}^b < \kappa < \underline{x}^a \leq \bar{x}^a < x_H$.*

(1) For sellers of type $x \in [\bar{x}^b, \kappa]$ and buyers of type $x \in [\kappa, \underline{x}^a]$, trade breaks down;

(2) for sellers of type $x \in [x_L, \underline{x}^b]$ and buyers of type $x \in [\bar{x}^a, x_H]$, trade is carried out without delay; but the bid price for sellers of type $x \in [x_L, \underline{x}^b]$ is the lowest in the market and the ask price for buyers of typex $\in [\bar{x}^a, x_H]$ is the highest;

(3) if there exists an equilibrium where q_{ax} and q_{bx} are continuous functions in x, then $\underline{x}^b < \bar{x}^b$ and $\underline{x}^a < \bar{x}^a$, trade for sellers of type $x \in (\underline{x}^b, \bar{x}^b)$ and buyers of type $x \in (\underline{x}^a, \bar{x}^a)$ is delayed but trade takes place at intermediate level of bid/ask spread.

The proposition is derived from Theorem 1. The proof is at the end of Section A.5 of the Appendix. It is a unique feature of the model that there might coexist three regions categorized by the surplus from trade. The literature on adverse selection with competitive search friction, initiated by Guerrieri, Shimer and Wright (2010)[18], only shows the second type of distortion, where all trade is delayed. Intuitively, this is because the reservation value of the dealer when he offers the screening contract is strictly above zero. So, trade always break down for those traders whose gain from trade even under complete information is too small. As in a static screening problem where type distribution is not endogenous, trade

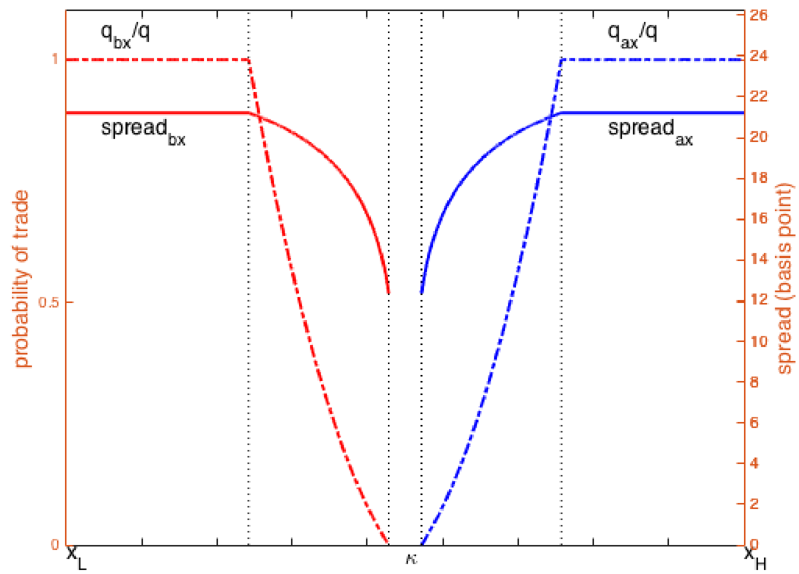

Figure 1: Equilibrium contract menus.

either breaks down or is not delayed at all. These features of equilibrium contract menus are illustrated in Figure 1. The bid or ask spreads, defined as $\frac{|p_{ix}-P|}{P}$, $i \in \{b, a\}$, are presented in basis points in Figure 1.

The parameter values used in the numerical example to generate the figure are calibrated to target empirical findings from municipal bonds market. I set the annual discount rate at 0.05 following the numerical example in Duffie, Garleanu and Pedersen (2005)[9] (DGP). The parameter values for relationship stability, γ, and relationship formation, α, target the municipal bond market. Li and Schuerhoff (2014)[29] study the network among financial institutions. Dealers in my model have a different meaning from those in Li and Schuerhoff (2014)[29]. They call all financial institutions dealers. I treat periphery dealers as traders in my model and core dealers in their paper as dealers in my model. Li and Schuerhoff (2014)[29] found that in the municipal bond market, the probability that two dealers who traded last month also trade this month is on average 65%, which implies an annualized arrival rate of the breakup shock of $\gamma = 0.35 \times 12 = 4.2$.[14,15] They also found that if two dealers did not trade in the previous month, there is on average an 85% chance that they do not trade with each other in the current month. Since traders matched with a dealer do not seek a second relationship, this number implies an annualized arrival rate of a trading relationship of $\alpha = 3.15$.[16] I assume that it takes at least two days for a dealer and trader to finish a transaction. So, $q = 182$, which is 7 times as large as the trading frequency used in DGP. In DGP, the flow utility from dividend consumption of high preference type traders is 10 times higher than that of low preference type traders. Following the specification, I assume the preference is distributed uniformly on $[.1, 1]$. Since I do not model trading frictions between dealers, I target the average turnover rate for trade between customers and dealers, defined as the aggregate trading volume between customers and dealers divided by the asset supply. Lester, Rocheteau and Weill (15)[28] offer an estimate of turnover for municipal bonds, based on the volume of dealer-to-customer data (i.e. not including dealer-to-dealer trades) of Green, Hollifield and Schueroff (07)[16], and aggregate bond holding data in the Flow of Funds. This turnover number is around 56%. The arrival rate of the preference shock is set at 1.525 to match the turnover rate number for trade between traders and dealers. These calibrated parameter values are summarized in Table 1.

If traders have direct access to the interdealer market, it is easy to show that there is no trading delay in equilibrium and the equilibrium allocation is efficient and is the same as in an environment with complete information. The following Corollary summarizes the effect of the endogenous liquidity distortion on the aggregate allocation.

Corollary 1. *Compared to the equilibrium allocation in an environment where traders have*

[14] We think of periphery dealers in the data as a collection of traders. Then, by the law of large numbers, there will be trade between periphery dealers and core dealers whenever they are matched.

[15] The trading relationship in the overnight interbank lending market is more persistent. Afonso, Kovner and Schoar (2014)[1] found that the monthly autocorrelation of volume share of lenders is about 91%.

[16] Only a fraction, $\frac{\alpha}{\alpha+\gamma}$, of traders are seeking trading relationships. Among these traders, the arrival rate of a trading relationship is α. So, α is approximately solved by $0.15 \times 12 = \frac{\gamma}{\alpha+\gamma} \times \alpha$, which implies that $\alpha = 3.15$.

Preference			Matching		
shock arrival rate	δ	1.525	meeting rate	α	3.15
upper bound	x_H	1	breakup rate	γ	4.2
lower bound	x_L	.1	Market		
discount rate	r	0.05	asset supply	A	0.5
distribution function	$G(x)$	$\frac{x-x_L}{x_H-x_L}$	trading frequency	q	182

Table 1: Parameter values used the numerical example.

direct access to the interdealer market, the equilibrium has the following qualitative properties: bid-ask spread arises; delay in trade arises; asset allocation is worse; aggregate trading flow is lower.

From Corollary 1, we can see that the equilibrium allocation of liquidity is not efficient. Figure 2 illustrates the inefficiency. The efficient allocation is calculated by assuming that all buyers and sellers trade at rate $q = 182$. The allocation in DGP is calculated by assuming that all trade takes place at rate $q = 26$, the value used in the numerical example in DGP.[17] Quantitatively, the allocation in DGP is uniformly close to the efficient allocation for traders of all types. In contrast, allocation efficiency is type dependent in this model. The equilibrium allocation in this model is the same as the autarky allocation for traders with small gain, whose trade breaks down. For traders with intermediate gain, trade is delayed. The allocation for them is in between the efficient allocation and the autarky allocation. For traders with large gain, the allocation is the same as the efficient allocation, which is better than the DGP allocation. But quantitatively, the additional improvement appears small. This numerical exercise shows that quantitatively, the model could predict more severe asset misallocation. This is because the asset misallocation for traders with small or intermediate

[17] To facilitate the comparison, what I call DGP in the exercise only allows trade between the dealer and those traders in long-term relationships with the dealer, just that conditional on being matched together, they act as in DGP, trading at frequency $q = 26$.

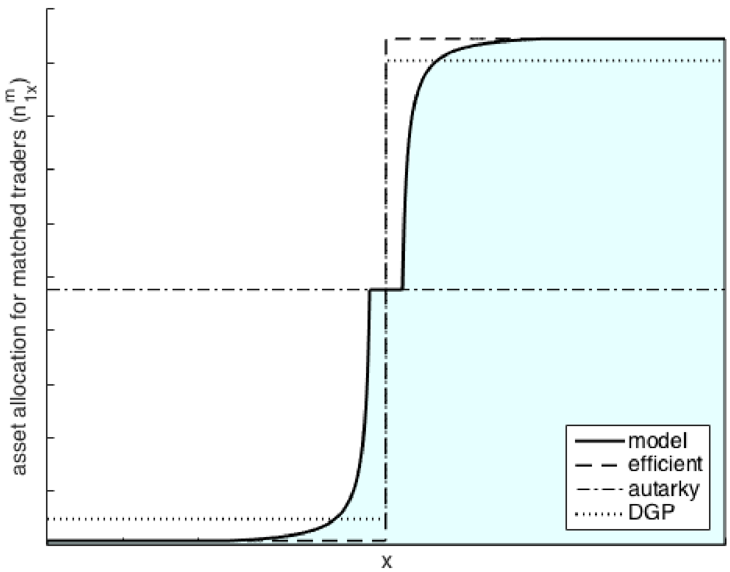

Figure 2: Equilibrium asset allocation.

gain could be a lot more severe than that in DGP.

4.2 Comparative Statics in the Numerical Example

In this section, I study the comparative statics in the calibrated numerical example. I focus a subset of parameters to be studied in the next section, all related to the market structure. The comparative statics act as a horse race among these parameters in explaining the liquidity disruption in the OTC market during the recent financial crisis.

In Table 2, I first report measures of liquidity in the benchmark which uses calibrated parameter values reported in Table 1. The measures of liquidity include the turnover rate, the probability of trade, conditional on traders being matched with a dealer, receiving a trading opportunity and having a positive gain from trade, the average, maximum and minimum spread among all trades. While the probability of trade is not directly observable in the data, it illustrates the response of dealers' strategic delay to changes in the market structure.

In the benchmark, the annualized turnover rate is 56%, the probability of trade is 0.15,

	Benchmark	Breakup Rate +100%	Matching Rate -50%	Trading Speed -50%
Turnover Rate (%)	56	-7.9%	-22.8%	-4.8%
Probability of Trade	.15	24.6%	0.2%	31.7%
Average Spread (bp)	20	12.0%	21.7%	36.9%
Maximum Spread (bp)	21	11.8%	21.7%	36.4%
Minimum Spread (bp)	12	17.0%	22.1%	44.6%
Welfare		-1.74%	-5.30%	-0.54%

Table 2: Comparative Statics in the Numerical Example.

average spread is 20 basis points, the maximum spread is 21 basis points and the minimum spread is 12 basis points. To derive clearer welfare implications, I also compute the welfare measure, W_t, defined as the total flow utility investors, matched and unmatched.

$$W_t \equiv \int_t^\infty e^{-r(s-t)}(n^m_{1xs} + n^u_{1xs})xdx.$$

Then, I study the relative changes in these liquidity measures when the fundamentals of the model change. Fundamental changes I study include increasing breakup rate, decreasing in matching rate and decrease in trading speed. The breakup rate represents stability of the trader-dealer relationship. The matching rate represents frictions in relationship formation. Trading speed represents trading frictions in DGP.[18]

The stability of trading relationship has a small impact on the efficiency of asset allocation. Although the breakup rate increases by 100%, the asset turnover rate only decreases by 7.9%, indicating that asset misallocation does not increase dramatically. This is because when expecting less stable relationships, dealers choose to screen traders by charging a higher spread rather than delaying trade. The probability of trade increases by 24.6% percent, offsetting the decrease in the turnover rate because of the higher breakup rate. The average spread, meanwhile, increases on average by 12.0%. When dealers delay less, the spread they

[18] See footnote 17.

charge becomes less dispersed. In this case, the minimum spread increases by 17.0%, 5.2% more than the increase in the maximum spread. The decrease in turnover rate is reflected in the decrease in the welfare measure, derive from their asset holding. The welfare measure decreases by 1.74%.

Decreasing the matching rate has a stronger effect on welfare. With a 50% decrease in matching rate, the turnover rate decreases by 22.8% and the welfare measure decreases by 5.30%. This is because as the matching rate decreases, dealers do not respond by increasing the trading speed. The probability of trade increases only by .2%. Meanwhile, the average spread they charge is 21.7% higher. Intuitively, this is because the strategic interaction between dealers becomes weaker when the matching rate is low, mitigating dynamic competition between dealers. So, they charge a higher spread while still strategically delaying trade.

Similar to the effect of relationship stability, dealers respond by reducing strategic delay when the trading speed limit, q, decreases. When q decreases by 50%, the probability trade increases by 31.7%. Dealers screen traders by charging higher spread instead of delaying trade. The spread increases on average by 36.9%.

While decreasing matching rate and decreasing trading speed have similar effects on equilibrium spread, they have very different effect on welfare. Asset turnover rate only decreases by 4.8% and the welfare measure decreases by .54% when trading speed decreases by 50%. Turnover rate drops by 4.8 times as much and the measure decreases by 5.3% when matching rate decreases by 50%. This shows it is important to distinguish frictions in relationship formation from limits on trading speed.

Empirically, Di Maggio, Kermani and Song (2016)[8] found that during the recent financial crisis, core dealers charge higher markups to their clients, without increasing liquidity provision. This is more consistent with increasing the matching friction in my model.

4.3 Comparative Statics in Analytical Examples

I study two analytical examples in this section, to analyze comparative statics in the numerical exercise and derive further comparative statics. The solutions to the two examples are in Section A.7 and A.9 of the Appendix.

Example 1. Permanent Relationship

If we assume that dealer-trader relationships are permanent, $\gamma = 0$, the model can be solved in closed form. Assume also that the distribution satisfies the following property: $g(x)(x-x_M)^2$ is increasing in x and $\frac{g(x)}{1-G(x)}$ is increasing in x when $x \geq x_M$.[19]

$$q_{ax} = \frac{g(x)}{g(\bar{x}^a)} \left(\frac{x-x_M}{\bar{x}^a - x_M}\right)^2 (\delta+q) - \delta, \forall x \in [\underline{x}^a, \bar{x}^a], \quad (31)$$

$$\bar{x}^a \; : \; 1 = \frac{g(\bar{x}^a)\,(\bar{x}^a - x_M)}{1-G(\bar{x}^a)}, \quad (32)$$

$$\underline{x}^a \; : \; \left(\frac{\underline{x}^a - x_M}{\bar{x}^a - x_M}\right)^2 \frac{g(\underline{x}^a)}{g(\bar{x}^a)} = \frac{\delta}{\delta+q}, \quad (33)$$

and q_{bx}, $\bar{x}^b$, $\underline{x}^b$ can be derived in the same fashion.

In this equilibrium, the qualitative properties in Proposition 1 and Corollary 1 hold. And we can also solve for bid or ask spreads.[20] Qualitatively, the spread is higher when there is more delay. So, we focus on delay in comparative statics.

Example 2. Discrete Distribution

Assume that the preference type follows a discrete distribution on $\{x_L, y, x_M, z, x_H\}$, with $x_L < y < x_M < z < x_H$. Assume that $\Pr(x = x_L) = \Pr(x = x_H) = \pi_1$, $\Pr(x = y) = \Pr(x = z) = \pi_2$. $A = \frac{1}{2}$. $x_M = \frac{1}{2}(x_L + x_H)$, $x_M - y = z - x_M$, and that agents are very patient, $r \ll \delta$.

[19] The characterization of bid contracts is symmetric and is omitted to economize the presentation.

[20] For example, if agents are patient in the sense that $r \ll 1$, $p_{ax} - P = \frac{\underline{x}^a - x_M}{\delta} + \int_{\underline{x}^a}^{x} \left[\frac{g(x)(x-x_M)^2}{g(s)(s-x_M)^2} - 1\right] ds, \forall x \in [\underline{x}^a, \bar{x}^a]$.

In this example, $q_{ax_H} = q_{bx_L} = q$, $q_{ax_M} = q_{bx_M} = 0$. Restricting our attention to parameter values such that $q_{az} = q_{by} \in (0, q)$, we have

$$q_{az} = \left[\phi \frac{(\alpha + \delta)\left(1 + \frac{q}{\delta}\right) + \gamma}{\alpha + \delta + \gamma} - 1 \right] (\delta + \gamma) \, . \tag{34}$$

where $\phi = \frac{z - x_M}{x_H - z} \frac{\pi_2}{\pi_1}$. The participation constraints of intermediate types are binding in equilibrium. So, the spread for intermediate types does not depend on equilibrium delays and is equal to the surplus from trade. The spread for type x_L and x_H is decreasing in the delay facing intermediate types. This analytical example is particularly useful to more completely characterize the general equilibrium effect of the frictions in forming and maintaining long-term relationships.

4.3.1 Frictions Related to the Long-term Relationship

The general equilibrium effects of the frictions in forming and maintaining long-term relationships are summarized in Corollary 2 and Corollary 3. The two corollaries are derived from equation (34).

Corollary 2. *Under the conditions in Example 2, a higher matching friction to form new relationships (lower α) corresponds to more equilibrium strategic delay in trade.*

The intuition behind Corollary 2 is the following. With a high matching friction, it takes a long time for an unmatched trader to find a dealer. Then, the equilibrium asset allocation to unmatched traders is closer to the autarky allocation and is less affected by dealers' screening strategies, as is evident from the law of motion of traders, equation (13). Therefore, an unmatched trader is more likely to have large gain from trade in this case. This gives dealers more incentive to screen those traders by delaying trade with other traders with smaller gain from trade. And the competition between dealers is less intense. In terms of efficiency, Corollary 2 implies that the matching friction can be amplified by strategic delay of dealers. This echoes the findings in the numerical example, where when matching friction increases, dealers do not significantly reduce strategic delay, leading to a big decrease in turnover rate and allocation efficiency.

The following Corollary shows that findings in the numerical example about the stability of trading relationship holds only when the matching friction is small.

Corollary 3. *Under the conditions in Example 2, the effect of relationship breakup on equilibrium strategic delay is ambiguous. If the matching friction is negligible ($\alpha \gg \max(\delta, \gamma)$), increasing the breakup rate decreases delay. If the matching friction is large ($\alpha \ll \delta$), increasing the breakup rate increases delay.*

The proof for the corollary is in Section A.10 of the Appendix. When the matching friction is small, relationship breakup induces more dynamic competition among dealers. But when the matching friction is large, the strategic interaction is muted. Meanwhile, a higher breakup rate means that the profit of a dealer depends more on the inflow of traders from the unmatched population, whose gain from trade is large because of the large matching friction. This induces dealers to delay trade with traders with intermediate gain from trade so as to screen traders with large gain from trade. Corollary 3 implies that the stability of the long-term relationship with dealers is more valuable for traders in an economy with fewer financial intermediaries, not only because of the exogenous frictions but also because it reduces strategic delay by dealers. One could imagine that during financial crises, fewer dealers participate in the market. Then, the matching friction increases.[21] Corollary 3 implies that the stability of relationship is particularly valuable during those periods.

4.3.2 The Frequency of Preference Shocks

The effect of preference shocks on liquidity distortion is summarized in Corollary 4.

Corollary 4. *When the trader-dealer relationship is permanent, under the conditions of Example 1, equilibrium delay increases for all traders as the preference shock becomes more frequent. In particular, the measure of traders for whom trade breaks down increases ($\bar{x}^b$ decreases and $\underline{x}^a$ increases). Equilibrium delay in trade strictly increases for those traders facing delay in trade (buyers with preference type $x \in [\underline{x}^a, \bar{x}^a]$ and sellers with preference type $x \in [\underline{x}_b, \bar{x}^b]$).*

[21] See Lagos and Zhang (2015)[26] for a formal model about the effect of dealer entry on market liquidity.

The proof is in Section A.11 of the Appendix. According to the law of motion of traders matched with dealers, equation (13), the asset allocation is closer to autarky allocation when the preference shock is more frequent. So, there are more traders with large gain from trade, this induces dealers to delay trade with traders with intermediate gain so as to screen traders with large gain from trade. Corollary 4 shows that in turbulent times, liquidity distortion could be higher because of strategic delay of the dealer.

Corollary 5. *Under the conditions in Example 2, increasing the frequency of preference shocks also induces more delay.*

Corollary 5 shows that findings in Corollary 4 is robust to the general equilibrium effects induced by the matching friction and relationship breakups. The proof in Section A.12 of the Appendix.

4.3.3 The Dispersion of Private Valuation

The dispersion of the preference shock affects the size of bid/ask spreads and information rent enjoyed by traders. This dispersion varies across assets, depending on such features as the opacity of the market and the sophistication of the asset under study.

I use Example 1 to study the effect of valuation dispersion, assuming that the distribution function satisfies the following functional form: $g(x) = \frac{1}{C}|x - x_M|^\zeta$ with $\zeta > -1$ and $C = \frac{(x_H - x_L)^{\zeta+1}}{2^\zeta(\zeta+1)}$. This class of distribution functions includes the uniform distribution as a special case ($\zeta = 0$). And a distribution function with a higher ζ has a thicker tail and higher dispersion. Corollary 6 shows that dispersed private valuation induces more delay.

Corollary 6. *Under the conditions in Example 2 and the conditions in Example 1 with $g(x) = \frac{1}{C}|x - x_M|^\zeta$ ($\zeta > -1$), increasing the dispersion of preference shocks increases trading delay for traders of all preference types.*

According to Corollary 6, then the dispersion increases, both the spread and delay in trade increase for traders of all preference types. But because there are more traders with large gain from trade in this case, the aggregate trading flow may not decrease.

Duffie, Malamud and Manso (2014)[10] shows that the opacity about heterogeneity in the asset quality also has important implications on trading activities in the market, and that the effect of opacity on price dispersion is not just driven by heterogeneity in asset quality but also by the dispersion in private valuation. Edwards, Harris and Piwowar (2007)[11] studies empirically how the opacity of the OTC markets of corporate bonds affects the transaction cost. They find that the transaction cost is linked to the opacity of the OTC market. This implies that the private information about heterogeneous preferences could also be an important source of distortion when the market structure is opaque.

5 Implementation by Random Bid-Ask Quotes

In this section, I show that in an environment where dealers are restricted to trading without delay, the optimal contract menus can still be implemented. Imagine that a dealer hires a continuum of employees, each holding a telephone line. Each trader is assigned an employee to handle his trade. Employees are busy most of the time doing other things such as handling trade for other customers. So, the trader needs to wait for her turn, which arrives with Poisson rate q. When her turn comes, she gets a random draw from some distribution of bid/ask quotes decided by the dealer. And the trade will be carried out without delay if the trader accepts the quote. Otherwise, the trader waits for her next turn to draw a quote from the distribution of bid/ask quotes specified by the dealer at that moment.

So far, it is sufficient to look at ex ante IR constraint for a contract menu to be feasible, because the quoted price is deterministic. In this section, the trading mechanism is implemented through price lotteries. So, it must also satisfy ex post IR constraints. That is, for contract menus $\mathbb{M}_{ix} = \{(q_{ix}, \tilde{p}_{ix})\}_{\forall x \in \mathbb{X}}, \forall i \in \{a, b\}$ to be feasible, not only should the expected price $E\tilde{p}_{ix}$ satisfy traders' ex ante IR constraint,

$$E\tilde{p}_{ax} \leq d_x \leq E\tilde{p}_{bx},$$

but the realization of the price $\tilde{p}_{ix}$ should also satisfy traders' ex post IR constraint,

$$\tilde{p}_{ax} \leq d_x \leq \tilde{p}_{bx}. \tag{35}$$

When dealers quote randomly, the strategy of traders can be characterized by their reservation value, like in a random search model with price dispersion. Buyers will buy if and only if the ask price is below their reservation value. Sellers will sell if and only if the bid price is above their reservation value. The reservation value is a function of traders' preference type. For example, buyers with a high preference type have a high reservation value. Therefore, they are more likely to accept an ask quote. Trade for those with large gain from trade is less likely to be delayed but the transaction cost is higher for them. This intuition is in line with the optimal contract menus designed by the dealer.

I show that if the dealer chooses optimally the distributions of bid and ask quotes, the profit maximizing contract menus can be implemented. Denote the reservation value of traders with asset holding a and preference type x to be R_{ax}. Denote the distribution of ask quotes to be $F_a(\cdot)$ and the distribution of bid quotes to be $F_b(\cdot)$.

Lemma 4. *The equilibrium contract menus $\mathbb{M}_{ix} = \{(q_{ix}, \tilde{p}_{ix})\}_{\forall x \in \mathbb{X}}, \forall i \in \{a, b\}$, the price distributions $F_a(\cdot)$ and $F_b(\cdot)$ implementing the equilibrium contract menus satisfy the following equations,*

$$\frac{q_{ax}}{q} = F_a(R_{0x}),\ \frac{q_{bx}}{q} = 1 - F_b(R_{1x}), \tag{36}$$

$$\frac{q_{ax}}{q} E\tilde{p}_{ax} = \int_{R_{0\kappa}}^{R_{0x}} R dF_a(R),\ \frac{q_{bx}}{q} E\tilde{p}_{bx} = \int_{R_{1x}}^{R_{1\kappa}} R dF_b(R). \tag{37}$$

$$\tilde{p}_{ax} \sim \min\left\{1, \frac{F_a(\tilde{p}_{ax})}{F_a(R_{0x})}\right\}, \tilde{p}_{bx} \sim \max\left\{0, \frac{F_b(\tilde{p}_{ax}) - F_b(R_{1x})}{1 - F_b(R_{1x})}\right\} \tag{38}$$

Lemma 4 gives a necessary condition for the optimal contract menu to be feasible and implementable by "random search". But it is not sufficient because the reservation value solved this way must satisfy the indifference condition: $R_{0x} = d_x$ for buyers of type x and $R_{1x} = d_x$ for sellers of type x. The following lemma confirms these conditions.

Lemma 5. *Given the quote distributions implied by the optimal contract menus, R_{ax} is indeed the reservation price for traders with asset holding a and preference type x. $R_{ax} = d_x$.*

Lemma 4 and Lemma 5 together prove the following proposition.

Proposition 2. *The optimal mechanism can be implemented by dealers quoting randomly from price distributions $F_a(\cdot)$ and $F_b(\cdot)$ following equations (36) and (37).*

The price dispersion predicted by this implementation is greater than the price dispersion predicted by the contract menus. This is because all traders whose trade does not break down enjoy some information rent, $p_{ax} < d_x$ for all $x \geq \underline{x}^a$, and $p_{bx} > d_x$ for all $x \leq \bar{x}^b$. We also know that the support of the distribution of ask(bid) prices is $[d_{\underline{x}^a}, d_{\bar{x}^a}]$ ($[d_{\underline{x}^b}, d_{\bar{x}^b}]$). But the difference in the price implication between the direct trading mechanism and this implementation is not quantitatively large.

The random pricing strategy of the dealer may also arise from competition among dealers. This mechanism is proposed by Burdett and Judd (1983)[4] and applied to study price dispersion in opaque financial markets by Green (2007)[15]. This implementation shows that even if the dealer has monopsony power, the price dispersion may still emerge in equilibrium as a result of dealers dynamically screening investors with heterogeneous valuations over the asset. Therefore, the paper offers an alternative mechanism linking the opacity of the secondary market to price dispersion.

6 Conclusion

I study in an OTC market the effect of unobservable private valuation on market liquidity and asset allocation when dealers and traders are in long-term relationships. I solve the optimal trading mechanism contingent on the aggregate history of the traders matched with a dealer, summarized by the asset allocation. The dynamic screening behavior of the dealer provides an additional mechanism to account for the liquidity distortion and price dispersion in the market.

A Appendix

A.1 Proof of Lemma 1.

Proof. I will present only the proof for ask contracts. The argument for bid contracts is symmetric. By IC constraints,

$$q_{axt}(d_{xt} - p_{axt}) \geq q_{ayt}(d_{xt} - p_{ayt}) \tag{39}$$

$$q_{axt}(d_{yt} - p_{axt}) \leq q_{ayt}(d_{yt} - p_{ayt}). \tag{40}$$

Subtracting (40) from (39), together with the condition that $d_{xt} > d_{yt}$ implies that $q_{axt} \geq q_{ayt}$. (40) implies that $q_{axt}p_{axt} \geq (q_{axt} - q_{ayt})d_{yt} + q_{ayt}p_{ayt}$. Because $q_{axt} \geq q_{ayt}$, this means that $q_{axt}p_{axt} \geq q_{ayt}p_{ayt}$. Using (39) and $d_{xt} > d_{yt}$, we can derive $q_{axt}(d_{xt} - p_{axt}) \geq q_{ayt}(d_{yt} - p_{ayt})$.

Property 4 follows from property 1 and the no-round trip trading condition, $q_{axt}q_{bxt} = 0$ for all x.

□

A.2 Proof of Lemma 2

Lemma 6. *Given feasible contract menus, any solutions to (8) and (11) that satisfy conditions (12) are strictly increasing and Liptschitz continuous.*

Proof. Using the property of the stopping time, the solution to (8) that satisfies the transversality condition is a fixed point of the operator

$$\Gamma_{xt}(d) = \int_t^\infty e^{-(r+\delta+\gamma)(\tau-t) - \int_t^\tau (q_{axs}+q_{bxs})ds} \left[x + (\delta + (q_{ax\tau} + q_{bx\tau}) + \gamma)\, d_{x\tau} + \Delta_{x\tau}(d)\right] d\tau. \tag{41}$$

where

$$\Delta_{x\tau}(d) \equiv \delta \int (d_{\tilde{x}\tau} - d_{x\tau})\, dG(\tilde{x}) + q_{ax\tau}p_{ax\tau} + q_{bx\tau}p_{bx\tau} - (q_{ax\tau} + q_{bx\tau})\, d_{x\tau} + \gamma(w_{x\tau} - d_{x\tau}).$$

Assume that $d_{xt} = \Gamma_{xt}(d)$ is a fixed point that satisfies (8) and (11). Since the right hand side of (41) is absolutely continuous in time, d_{xt} inherits this property, and it follows from Lebesgue's differentiation theorem that

$$\dot{d}_{xt} = -x - \Delta_{xt}(d) + rd_{xt}$$

Using this equation together with the equation, $\int_t^T e^{-r(\tau-t)} d_{x\tau} d\tau = \frac{1}{r}\left(d_{xt} - d_{xT} e^{-r(T-t)}\right) + \frac{1}{r}\int_t^T e^{-r(\tau-t)} \dot{d}_{x\tau} d\tau$, we get

$$d_{xt} = \int_t^T e^{-r(\tau-t)}(x + \Delta_{x\tau}(d))d\tau + e^{-r(T-t)} d_{xT}. \tag{42}$$

$$= \int_t^\infty e^{-r(\tau-t)}(x + \Delta_{x\tau}(d))d\tau. \tag{43}$$

Since the right hand side of equation (42) is absolutely continuous in t, the fixed point d_{xt} inherits the property. Assume by contradiction that there exists $x > y$ such that $d_{xt} \leq d_{yt}$, for all t from t_0 to T, where $T \equiv \inf\{\tau : d_{x\tau} > d_{y\tau},\ \tau \geq t_0\}$, then

$$\begin{aligned} d_{xt} - d_{yt} = &\int_t^T e^{-r(\tau-t)} \{x - y + \gamma(w_{x\tau} - w_{y\tau} - (d_{x\tau} - d_{y\tau})) + \delta\left(d_{y\tau} - d_{x\tau}\right) \\ &+ \left[\left(q_{ax\tau}p_{ax\tau} + q_{bx\tau}p_{bx\tau} - \left(q_{ax\tau} + q_{bx\tau}\right) d_{x\tau}\right)\right. \\ &\left.\left. - \left(q_{ay\tau}p_{ay\tau} + q_{by\tau}p_{by\tau} - \left(q_{ay\tau} + q_{by\tau}\right) d_{y\tau}\right)\right]\right\} d\tau + e^{-r(T-t)}\left(d_{xT} - d_{yT}\right). \end{aligned}$$

Next, we will show that $d_{xt} \leq d_{yt}$ for all t from t_0 to T implies

$$\int_t^T e^{-r(\tau-t)}(w_{x\tau} - w_{y\tau} - (d_{x\tau} - d_{y\tau}))d\tau \geq 0, \forall t \in [t_0, T],$$

$$\begin{aligned} &\left(q_{axt}p_{axt} + q_{bxt}p_{bxt} - \left(q_{axt} + q_{bxt}\right) d_{xt}\right) \\ &- \left(q_{ayt}p_{ayt} + q_{byt}p_{byt} - \left(q_{ayt} + q_{byt}\right) d_{yt}\right) \geq 0, \forall t \in [t_0, T]. \end{aligned}$$

Because of the transversality condition, this equation holds regardless of whether T is finite or infinite. Using the property of the stopping time, we can rewrite (11) as

$$\begin{aligned} w_{xt} = &\int_t^H e^{-(\alpha+\delta)(\tau-t)} \left[(\alpha+\delta)x \int_t^\tau e^{-r(s-t)} ds + e^{-r(\tau-t)}\left(\alpha d_{x\tau} + \delta \int w_{\tilde{x}\tau} dG(\tilde{x})\right)\right] d\tau \\ &+ e^{-(\alpha+\delta)(H-t)}\left[x \int_t^H e^{-r(s-t)} ds + e^{-r(H-t)} w_{xH}\right] \end{aligned}$$

for any $H > t$. By transversality conditions, if $T < \infty$,

$$\begin{aligned} w_{xT} - w_{yT} = &\int_T^\infty e^{-(\alpha+\delta)(\tau-T)} \left[(\alpha+\delta)(x-y)\int_T^\tau e^{-r(s-T)} ds + e^{-r(\tau-t)} \alpha\left(d_{x\tau} - d_{y\tau}\right)\right] d\tau \\ &\geq 0. \end{aligned}$$

The inequality follows from the definition of T. Then, for any $t \in [t_0, T]$,

$$\begin{aligned} w_{x\tau} - w_{y\tau} &= \int_\tau^T e^{-(\alpha+\delta)(s-\tau)} \left[(\alpha+\delta)(x-y)\int_\tau^s e^{-r(v-\tau)}dv + e^{-r(s-\tau)}\alpha\left(d_{xs} - d_{ys}\right)\right] ds \\ &\quad + e^{-(\alpha+\delta)(T-\tau)} \left[(x-y)\int_\tau^T e^{-r(s-\tau)}ds + e^{-r(T-\tau)}\left(w_{xT} - w_{yT}\right)\right] \\ &\geq \int_\tau^T e^{-(r+\alpha+\delta)(s-\tau)}\alpha\left(d_{xs} - d_{ys}\right) ds. \end{aligned}$$

Therefore,

$$\begin{aligned} &\int_t^T e^{-r(\tau-t)}(w_{x\tau} - w_{y\tau} - (d_{x\tau} - d_{y\tau}))d\tau \\ &\qquad \geq \int_t^T e^{-r(\tau-t)} \left(\int_\tau^T e^{-(r+\alpha+\delta)(s-\tau)}\alpha\left(d_{xs} - d_{ys}\right) ds - (d_{x\tau} - d_{y\tau})\right) d\tau \\ &\qquad = \int_t^T e^{-r(\tau-t)} \left(-1 + \int_t^\tau e^{-(r+\alpha+\delta)(\tau-v)}\alpha dv\right)(d_{x\tau} - d_{y\tau})d\tau \\ &\qquad = \int_t^T e^{-r(\tau-t)} \left(-1 + \frac{\alpha}{r+\alpha+\delta}(1 - e^{-(r+\alpha+\delta)(\tau-t)})\right)(d_{x\tau} - d_{y\tau})d\tau \\ &\qquad \geq 0. \end{aligned}$$

From Lemma 1,

$$q_{bxt}\left(p_{bxt} - d_{xt}\right) - q_{byt}\left(p_{byt} - d_{yt}\right) \geq 0, \tag{44}$$

$$q_{axt}\left(p_{axt} - d_{xt}\right) - q_{ayt}\left(p_{ayt} - d_{yt}\right) \geq 0. \tag{45}$$

Adding up (44) and (45) implies that

$$q_{axt}p_{axt} + q_{bxt}p_{bxt} - \left(q_{axt} + q_{bxt}\right)d_{xt} \geq q_{ayt}p_{ayt} + q_{byt}p_{byt} - \left(q_{ayt} + q_{byt}\right)d_{yt}.$$

If T is finite, by continuity of d_{xt} in t, $d_{xT} - d_{yT} = 0$. If T is infinite, by the transversality condition, $\lim_{T\to\infty} e^{-r(T-t)}\left(d_{xT} - d_{yT}\right) = 0$. Together, these imply that $d_{xt} - d_{yt} > 0$ for $t \in [t_0, T)$, which delivers the contradiction.

Following similar arguments, we can show that

$$w_{xt} = \int_t^\infty e^{-(r+\delta+\alpha)(\tau-t)}\left[x + \delta\int w_{\tilde{x}\tau}dG(\tilde{x}) + \alpha d_{x\tau}\right] d\tau \tag{46}$$

is strictly increasing in x.

□

Proof of Lemma 2 follows:

Proof. Denote the space of uniformly bounded, measurable functions from $[x_L, x_H] \times [0, \infty)$ to $\mathbb{R}_+$ equipped with supnorm to be $\mathcal{D}$. Clearly, Γ maps $\mathcal{D}$ into itself. Given $d_{x\tau}$, we can rewrite equation (11) as

$$w_{xt} \equiv \Pi_{xt}(d) = \int_t^\infty e^{-(r+\delta+\alpha)(\tau-t)} \left[x + \delta \int w_{\tilde{x}\tau} dG(\tilde{x}) + \alpha d_{x\tau} \right] d\tau. \tag{47}$$

$$\begin{aligned} \Pi_{x\cdot}(f) - \Pi_{x\cdot}(g) &= \int_t^\infty e^{-(r+\delta+\alpha)(\tau-t)} \alpha \left(f_{x\tau} - g_{x\tau} \right) d\tau \\ &\leq \frac{\alpha}{r+\delta+\alpha} \left| f - g \right| < \left| f - g \right|, \end{aligned}$$

where $|\cdot|$ denotes the supnorm. So, Π is a contraction mapping of uniformly bounded measurable functions from $[0, \infty]$ to $\mathbb{R}_+$ equipped with supnorm. From the contraction mapping theorem, there is a unique solution of w_{xt} corresponding to d_{xt}.

Rewrite equation (8) as

$$\begin{aligned} \Gamma_{xt}(d) = \int_t^\infty e^{-(r+\delta+\gamma+q)(\tau-t)} & \left[x + \delta \int d_{\tilde{x}\tau} dG(\tilde{x}) \right. \\ & \left. + \gamma \Pi_{x\tau}(d) + (q - q_{ax\tau} - q_{bx\tau}) d_{x\tau} + q_{ax\tau} p_{ax\tau} + q_{bx\tau} p_{bx\tau} \right] d\tau. \end{aligned}$$

To show that Γ is a contraction mapping given the sequence of contract menus, denote the contract type x trader chooses from the menu, given his reservation value $f(x)$, to be $\left(q^f_{am(x)\cdot}, p^f_{am(x)\cdot}, q^f_{bm(x)\cdot}, p^f_{bm(x)\cdot} \right)$,

$$\begin{aligned} \Gamma_{x\cdot}(f) - \Gamma_{x\cdot}(g) = \int_t^\infty e^{-(r+\gamma+\delta+q)\tau} & \left[\delta \int \left(f_{\tilde{x}\tau} - g_{\tilde{x}\tau} \right) dG(\tilde{x}) \right. \\ & + \gamma (\Pi_{x\tau}(f) - \Pi_{x\tau}(g)) \\ & - \left(q^g_{am(x)\tau} p^g_{am(x)\tau} + q^g_{bm(x)\tau} p^g_{bm(x)\tau} + \left(q - \left(q^g_{am(x)\tau} + q^g_{bm(x)\tau} \right) \right) g_{x\tau} \right) \\ & \left. + \left(q^f_{am(x)\tau} p^f_{am(x)\tau} + q^f_{bm(x)\tau} p^f_{bm(x)\tau} + \left(q - \left(q^f_{am(x)\tau} + q^f_{bm(x)\tau} \right) \right) f_{x\tau} \right) \right] d\tau. \end{aligned}$$

Denote

$$\begin{aligned} A_{x\tau} = & - \left(q^g_{am(x)\tau} p^g_{am(x)\tau} + q^g_{bm(x)\tau} p^g_{bm(x)\tau} + \left(q - \left(q^g_{am(x)\tau} + q^g_{bm(x)\tau} \right) \right) g_{x\tau} \right) \\ & + \left(q^f_{am(x)\tau} p^f_{am(x)\tau} + q^f_{bm(x)\tau} p^f_{bm(x)\tau} + \left(q - \left(q^f_{am(x)\tau} + q^f_{bm(x)\tau} \right) \right) f_{x\tau} \right). \end{aligned}$$

According to Lemma 1, $q^f_{am(x)\tau}p^f_{am(x)\tau} + q^f_{bm(x)\tau}p^f_{bm(x)\tau} - \left(q^f_{am(x)\tau} + q^f_{bm(x)\tau}\right) f_{x\tau}$ is decreasing in f_{xt}. So, for $f_{x\tau} > g_{x\tau}$, $A_{x\tau} < q(f_{x\tau} - g_{x\tau})$. And

$$
\begin{aligned}
A_{x\tau} > &- \left(q^g_{am(x)\tau}p^g_{am(x)\tau} + q^g_{bm(x)\tau}p^g_{bm(x)\tau} + \left(q - \left(q^g_{am(x)\tau} + q^g_{bm(x)\tau}\right)\right) f_{x\tau}\right) \\
&+ \left(q^f_{am(x)\tau}p^f_{am(x)\tau} + q^f_{bm(x)\tau}p^f_{bm(x)\tau} + \left(q - \left(q^f_{am(x)\tau} + q^f_{bm(x)\tau}\right)\right) g_{x\tau}\right).
\end{aligned}
$$

Under the no-round-trip-trading constraint, that $q_{ax\tau}q_{bx\tau} = 0$ for all x, if a trader with reservation value $g_{x\tau}$ chooses a bid contract,

$$
\begin{aligned}
&q^f_{am(x)\tau}p^f_{am(x)\tau} + q^f_{bm(x)\tau}p^f_{bm(x)\tau} - \left(q^f_{am(x)\tau} + q^f_{bm(x)\tau}\right) g_{x\tau} \\
&> q^g_{am(x)\tau}p^g_{am(x)\tau} + q^g_{bm(x)\tau}p^g_{bm(x)\tau} - \left(q^g_{am(x)\tau} + q^g_{bm(x)\tau}\right) g_{x\tau}.
\end{aligned}
$$

Therefore, $A_{x\tau} > -q(f_{x\tau} - g_{x\tau})$. If she chooses an ask contract, since $f_{x\tau} > g_{x\tau}$, a trader with reservation value $f_{x\tau}$ also chooses an ask contract,

$$
\begin{aligned}
&q^g_{am(x)\tau}p^g_{am(x)\tau} + q^g_{bm(x)\tau}p^g_{bm(x)\tau} + \left(q - \left(q^g_{am(x)\tau} + q^g_{bm(x)\tau}\right)\right) f_{x\tau} \\
&> q^f_{am(x)\tau}p^f_{am(x)\tau} + q^f_{bm(x)\tau}p^f_{bm(x)\tau} + \left(q - \left(q^f_{am(x)\tau} + q^f_{bm(x)\tau}\right)\right) f_{x\tau}.
\end{aligned}
$$

Therefore, $A_{x\tau} > -q(f_{x\tau} - g_{x\tau})$, and

$$
|A_{x\tau}| < q\,|f_{x\tau} - g_{x\tau}|\,. \tag{48}
$$

Together, these imply

$$
|\Gamma(f) - \Gamma(g)| \leq \frac{\gamma + \delta + q}{r + \gamma + \delta + q}\,|f - g|\,.
$$

So, Γ is a contraction mapping. The uniqueness of d then follows from the contraction mapping theorem.

For any $d_{x\tau}$ in $\mathcal{D}$, there exists a unique function $\int w_{x\cdot}dG(x)$ and therefore a unique reservation value function $w_{x\tau}$ in $\mathcal{D}$.

Given the monotonicity of d_{xt}, for any x and y such that $x_H \geq x > y \geq x_L$, any solution to the mapping, $\Gamma(d)$, satisfies

$$
\begin{aligned}
d_{xt} - d_{yt} = &\int_t^\infty e^{-(r+\gamma+\delta+q)\tau}\left[x - y + \gamma(w_{x\tau} - w_{y\tau})\right. \\
&- \left(q_{ax\tau}p_{ax\tau} + q_{bx\tau}p_{bx\tau} + (q - (q_{ax\tau} + q_{bx\tau}))\,d_{x\tau}\right) \\
&\left.+ \left(q_{ay\tau}p_{ax\tau} + q_{by\tau}p_{by\tau} + (q - (q_{ay\tau} + q_{by\tau}))\,d_{y\tau}\right)\right] d\tau \\
\leq &\int_t^\infty e^{-(r+\gamma+\delta+q)\tau}\left[x - y + \gamma(w_{x\tau} - w_{y\tau}) + q(d_{x\tau} - d_{y\tau})\right] d\tau.
\end{aligned} \tag{49}
$$

The inequality follows from Property 3 of Lemma 1, that $q_{axt}p_{axt} + q_{bxt}p_{bxt} - (q_{axt} + q_{bxt})\, d_{xt}$ is decreasing in d_{xt}.

$$w_{xt} - w_{yt} = \int_t^{\infty} e^{-(r+\delta+\alpha)(\tau-t)} \left[x + \alpha\left(d_{x\tau} - d_{y\tau}\right)\right] d\tau. \tag{50}$$

Let $\inf_{t\in[0,\infty)} d_{xt} - d_{yt} = k(x-y)$. From (50),

$$w_{xt} - w_{yt} < \frac{1+\alpha k}{r+\delta+\alpha}(x-y).$$

From (49),

$$d_{xt} - d_{yt} \leq \frac{1 + \gamma\frac{1+\alpha k}{r+\delta+\alpha} + qk}{r+\gamma+\delta+q}(x-y).$$

By definition of k,

$$k \leq \frac{1 + \gamma\frac{1+\alpha k}{r+\delta+\alpha} + qk}{r+\gamma+\delta+q}.$$

Therefore, k is a finite number and is independent of x and y and d_{xt} is Lipschitz continuous in x.

Since d_{xt} is also absolutely continuous in t, following Hugonnier, Lester and Weill (2014) [22] and the reference therein, d_{xt} is absolutely continuous.

Given d and feasible contract menus, $(M_{0xt}, M_{1xt}, U_{0xt}, U_{1xt})'$ is a fixed point of the following mapping,

$$\Xi_{xt}\left(\begin{bmatrix} M_0 \\ M_1 \\ U_0 \\ U_1 \end{bmatrix}\right) = \begin{bmatrix} \int_t^{\infty} e^{-(r+\delta+\gamma)(\tau-t)-\int_t^{\tau} q_{axs}ds}\left[x + \delta\int M_{0\tilde{x}\tau}dG(\tilde{x}) + \gamma U_{0x\tau} + q_{ax\tau}(d_{x\tau} - p_{ax\tau})\right] d\tau \\ \int_t^{\infty} e^{-(r+\delta+\gamma)(\tau-t)-\int_t^{\tau} q_{bxs}ds}\left[\delta\int M_{1\tilde{x}\tau}dG(\tilde{x}) + \gamma U_{1x\tau} + q_{bx\tau}(p_{bx\tau} - d_{x\tau})\right] d\tau \\ \int_t^{\infty} e^{-(r+\delta+\alpha)(\tau-t)}\left[x + \delta\int U_{0\tilde{x}\tau}dG(\tilde{x}) + \alpha M_{0x\tau}\right] d\tau \\ \int_t^{\infty} e^{-(r+\delta+\alpha)(\tau-t)}\left[\delta\int U_{1\tilde{x}\tau}dG(\tilde{x}) + \alpha M_{1x\tau}\right] d\tau \end{bmatrix}$$

The solution to the set of equations is a fixed point of the following mapping

$$\left|\Xi_{xt}(f) - \Xi_{xt}(g)\right| \leq \min\left\{\frac{\delta+\gamma}{r+\delta+\gamma}, \frac{\alpha+\delta}{r+\alpha+\delta}\right\}\left|f-g\right|.$$

Therefore, Ξ is a contraction mapping, the solution to the problem is unique given any reservation value functions d. $M_{1xt} - M_{0xt}$ corresponding to d as a fixed point of Γ also satisfies the equation for the reservation value function, d_{xt}. Therefore the solution of M_{0xt} and M_{1xt} given the fixed point of d is consistent with the reservation value function, $d_{xt} = M_{1xt} - M_{0xt}$. It is then straightforward to show that the solution of U_{1xt}, U_{0xt} satisfies $U_{1xt} - U_{0xt} = w_{xt} = \Pi_{xt}(d)$.

$\square$

A.3 Proof for Lemma 3

I only present the proof for the conditions for ask contracts here. The proof for the conditions for bid contracts is symmetric and is thus omitted.

First, from Lemma 2, d_{xt} is strictly increasing in x and is differentiable almost everywhere.

Necessity. The monotonicity of q_{axt} and q_{bxt} follows from Lemma 1.

Since $x \in \arg\max_m q_{amt}(d_{xt} - p_{amt})$ and from Lemma 2, d_x is differentiable almost everywhere, applying the Envelope Theorem where d_{xt} is differentiable, we have $q_{axt}(d_{xt} - p_{axt}) = \int_{\kappa_t}^{x} q_{ast} d'_{st} ds$, which means that $q_{axt} p_{axt} = q_{axt} d_{xt} - \int_{\kappa_t}^{x} q_{ast} d'_{st} ds$.

Sufficiency. It is optimal for a trader of type x to choose the contract with index x. Comparing her payoff from choosing contract x and contract m, we have,

$$
\begin{aligned}
& q_{axt}(d_{xt} - p_{axt}) - q_{amt}(d_{xt} - p_{amt}) \\
= & \int_{\kappa_t}^{x} q_{ast} d'_{st} ds - \int_{\kappa_t}^{m} q_{ast} d'_{st} ds - q_{amt}(d_{xt} - d_{mt}) \\
= & \int_{m}^{x} (q_{ast} - q_{amt}) d'_{st} ds \geq 0, \forall m \in [x_L, x_H].
\end{aligned}
$$

The first equality uses the condition that $q_{axt} p_{axt} = q_{axt} d_{xt} - \int_{\kappa_t}^{x} q_{ast} d'_{st} ds$. The inequality uses the condition that q_{amt} is weakly increasing in m. Q.E.D.

A.4 Dynamic Optimization of Dealers

Solutions satisfying the Maximum Principle. To solve for the dealer's screening problem, (23), we first write down the corresponding date-s present value Hamiltonian without the following constraints: $q_{ixs} \geq 0$, $q_{ixs} \leq q$, for all $x \in [0, q]$, $i = a, b$, and the monotonicity constraints for q_{ixs}.

The present value Hamiltonian is

$$
\mathcal{H}_{s,t} = \mathcal{D}_{s,t} + \int \lambda_{xs,t} f(x, q_{a \cdot s}, q_{b \cdot s}, n^m_{1 \cdot s}, s) dx, \forall s \geq t,
$$

where $\mathcal{D}_s$ follows (24), and

$$
f(x, q_{a\cdot}, q_{b\cdot}, n^m_{1\cdot s}, s) \equiv \alpha n^u_{1xs} - \delta n^m_{1xs} + \delta g(x) \int n^m_{1zs} dz - \gamma n^m_{1xs} - q_{bx} n^m_{1xs} + q_{ax}\left(\frac{\alpha}{\alpha + \gamma} g(x) - n^m_{1xs}\right).
$$

Define the derivative of $\mathcal{H}_s$ with respect to the value of function $u_{\cdot s} : [x_L, x_H] \to \mathbb{R}_+$ at y to be

$$\begin{aligned}\frac{\partial \mathcal{H}_{s,t}}{\partial u_{ys}} \equiv \int \delta(y-x)\frac{\partial}{\partial u_{ys}}\Big[& e^{-r(s-t)}(q_{axs}d_{xs} - \int_{\kappa_s}^{x} q_{aus}d'_{us}du - q_{axs}P_s)n^m_{0xs} \\ & + e^{-r(s-t)}(q_{bxs}P_s - q_{bxs}d_{xs} - \int_{x}^{\kappa_s} q_{bus}d'_{us}du)n^m_{1xs} \\ & +\lambda_{xs,t}f(x, q_{a\cdot s}, q_{b\cdot s}, n^m_{1\cdot s}, s)\Big]\, dx,\end{aligned} \tag{51}$$

where $\delta(y-x)$ is a Dirac delta function, λ_{xs} is the co-state variable for n^m_{1xs}.

Because any convex combination of two sequences of $\{q_{axs}, q_{bxs}\}_{s\geq t}$, that satisfy feasibility constraints is also feasible, the choice set for the control variables, trading frequencies, is convex. So, the optimization problem solving for q_{axs} and q_{bxs} is well-defined on the infinite-dimensional vector space. Then, using perturbation to q_{axs} and q_{bxs}, we can show that equations (52) and (53) hold. (54) follows from the maximum principle.

$$\frac{\partial \mathcal{H}_{s,t}}{\partial q_{axs}} = e^{-r(s-t)}\left[(d_{xs} - P_s)\, n^m_{0xs} - d'_{xs}\int_x^{x_H} n^m_{0us}du\right] + \lambda_{xs,t}n^m_{0xs} = 0, \tag{52}$$

$$\frac{\partial \mathcal{H}_{s,t}}{\partial q_{bxs}} = e^{-r(s-t)}\left[(P_s - d_{xs})\, n^m_{1xs} - d'_{xs}\int_{x_L}^{x} n^m_{1us}du\right] - \lambda_{xs,t}n^m_{1xs} = 0, \tag{53}$$

$$\begin{aligned}\frac{\partial \mathcal{H}_{s,t}}{\partial n^m_{1xs}} &= e^{-r(s-t)}\left(q_{bxs}(P_s - d_{xs}) - q_{axs}(d_{xs} - P_s) + \int_{x_L}^{x} q_{aus}d'_{us}du - \int_x^{x_H} q_{bus}d'_{us}du\right) \\ &- \lambda_{xs,t}(\delta+\gamma+q_{bxs}+q_{axs}) + \delta\int \lambda_{zs}g(z)dz = -\dot{\lambda}_{xs,t}.\end{aligned} \tag{54}$$

$\dot{\lambda}_{xs,t}$ refers to the derivative of $\lambda_{xs,t}$ with respect to s. In the steady state,

$$\dot{\lambda}_{xs,t} = -r\lambda_{xs,t}.$$

So, using Conjecture 1, for $x > \kappa$,

$$\lambda_{xs,t} = e^{-r(s-t)}\left[\frac{\int_\kappa^x (q_{au} - q_{ax})\, d'_u du}{r+\delta+\gamma+q_{ax}} + \delta\frac{\int \lambda_{zs}dG(z)}{r+\delta+\gamma+q_{ax}}\right]. \tag{55}$$

Similarly, for $x < \kappa$,

$$\lambda_{xs,t} = e^{-r(s-t)}\left[\frac{\int_x^\kappa (q_{bx} - q_{bu})\, d'_u du}{r+\delta+\gamma+q_{bx}} + \delta\frac{\int \lambda_{zs}dG(z)}{r+\delta+\gamma+q_{bx}}\right]. \tag{56}$$

Combining (55) and (56),

$$\int \lambda_{xs,t}dG(x) = e^{-r(s-t)}\frac{\int_\kappa^{x_H} \frac{\int_\kappa^x (q_{au}-q_{ax})d'_u du}{r+\delta+\gamma+q_{ax}}dG(x) + \int_{x_L}^\kappa \frac{\int_x^\kappa (q_{bx}-q_{bu})d'_u du}{r+\delta+\gamma+q_{bx}}dG(x)}{1 - \int \frac{\delta}{r+\delta+\gamma+q_{bx}+q_{ax}}dG(x)}.$$

In the symmetric equilibrium,

$$\int \lambda_{xs,t} dG(x) = 0.$$

Substitute (55) into (52), we have

$$\begin{aligned}
&\hat{v}_x = 0, \text{ where} \\
&\hat{v}_x = \frac{\int_\kappa^x (r+\delta+\gamma+q_{au}) d'_u du}{r+\delta+\gamma+q_{ax}} - \frac{\int_x^{x_H} n^m_{0u} du}{n^m_{0x}} d'_x.
\end{aligned}$$

We now show that C_x is continuously differentiable if $g(x)$ is continuous and therefore v_x is a continuous and increasing function in x. This is because $\hat{v}_x$ is continuous except at countable points of x where q_{ax} increases discontinuously. But $\hat{v}_x$ strictly decreases at those points because

$$\begin{aligned}
\frac{\partial \hat{v}_x}{\partial q_{ax}} = &- \frac{\int_\kappa^x (r+\delta+\gamma+q_{as}) \, d'_s ds}{r+\delta+\gamma+q_{ax}} \\
&- \frac{\int_x^{x_H} n^m_{0s} ds}{g(x) \frac{\alpha(\alpha+\delta+\gamma)}{\alpha+\gamma}(1-A)} \frac{r+\alpha+\delta+\gamma}{r+\delta} \frac{\frac{\alpha+\delta}{\delta} r + (\alpha+\delta+\gamma) \frac{r\alpha}{\delta(r+\delta)}}{\left(r+\delta+\alpha+\gamma+\frac{r+\delta+\alpha}{r+\delta} q_{ax}\right)^2} < 0.
\end{aligned}$$

Therefore, $\hat{v}_x$ is not convex at the neighborhood of those points. Specifically, denote any of those points as z. There exists an open neighborhood around z, denoted, $N(z)$, such that for any $\omega \in (0,1)$ $z_L, z_H \in N(z)$, $z_L < z_H$, such that $\omega z_L + (1-\omega) z_H = z$, $\omega \hat{C}_{z_L} + (1-\omega)\hat{C}_{z_H} < \hat{C}_z$, which implies that $C_z < \hat{C}_z$. Then, there exists a continuously differentiable function $\tilde{C}$ such that $C_x \leq \tilde{C}_x \leq \hat{C}_x$ for any x. Following Kirchheim and Kristensen (2001)[24], the convex envelope of a continuously differential function is continuously differentiable. Therefore, $v_x \equiv C'_x$ is a continuously increasing function.

Sufficiency for optimality. The maximum principle only provides the necessary condition for optimality. To verify that the conditions are sufficient, I adapt a method from Chapter 5.1.4 of Liberzon (2012)[30], which is about the sufficient condition for optimality of solutions to HJB equations. The key is that the optimality condition for trading frequencies is sufficient for dealers' subproblem, given any values of state and costate variables. Then, we can show that following the path of distributions corresponding to an alternative path of trading frequencies would lower dealer's expected profit.

Denote the date-t present value of a dealer's date-s flow profit to $\mathcal{D}_{s,t}(q_{a\cdot s}, q_{b\cdot s}, n^m_{1\cdot s})$, emphasizing that the value is a function of control and state variables. Notice that for all distributions, $n^m_{1\cdot s}$

and the corresponding co-state variable $\lambda_{xs,t}(n^m_{1\cdot s}, s)$, satisfying the maximum principle, the optimal trading frequencies, $q_{a\cdot s}, q_{b\cdot s}$, dominate any other feasible trading frequencies, $\tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}$. That is,

$$\begin{aligned}&\mathcal{D}_{s,t}(q_{a\cdot s}, q_{b\cdot s}, n^m_{1\cdot s}) + \int \lambda_{xs,t}(n^m_{1\cdot s}) f(x, q_{a\cdot s}, q_{b\cdot s}, n^m_{1\cdot s}, s) dx \\ \geq &\mathcal{D}_{s,t}(\tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}, n^m_{1\cdot s}) + \int \lambda_{xs,t}(n^m_{1\cdot s}) f(x, \tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}, n^m_{1\cdot s}, s) dx.\end{aligned} \tag{57}$$

Denote dealer's date-t expected payoff from date s on following the solution satisfying the maximum principle to be $D_{s,t}(n^m_{1\cdot s})$, all $s \geq t$. Denote dealer's date-t expected payoff from date s on following an alternative path of trading frequencies and the corresponding distributions, $\{\tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}\}_{s\geq t}$ and $\{\tilde{n}^m_{1\cdot s}\}_{s>t}$, given $\tilde{n}^m_{1\cdot t} = n^m_{1\cdot t}$, to be $\tilde{D}_{s,t}(n^m_{1\cdot s})$.

Dealer's expected payoff is solved by the following Lagrangian

$$\begin{aligned}D_{s,t}(n^m_{1\cdot s}) &= \int_s^\infty \left\{ \mathcal{D}_{\tau,t}(q_{a\cdot\tau}, q_{b\cdot\tau}, n^m_{1\cdot\tau}) + \int \lambda_{x\tau,t}(n^m_{1\cdot\tau}) \left[f(x, q_{a\cdot\tau}, q_{b\cdot\tau}, n^m_{1\cdot\tau}, \tau) - \dot{n}^m_{1x\tau} \right] dx \right\} d\tau \\ &= \int_s^\infty \left\{ \mathcal{D}_{\tau,t}(q_{a\cdot\tau}, q_{b\cdot\tau}, n^m_{1\cdot\tau}) + \int \left[\lambda_{x\tau,t}(n^m_{1\cdot\tau}) f(x, q_{a\cdot\tau}, q_{b\cdot\tau}, n^m_{1\cdot\tau}, \tau) + n^m_{1x\tau}\dot{\lambda}_{x\tau,t} \right] dx \right\} d\tau \\ &+ \int \lambda_{xs,t} n^m_{1xs} dx - \lim_{\tau\to\infty} \int \lambda_{x\tau,t} n^m_{1x\tau} dx.\end{aligned}$$

The second equality is derived using integration by parts. Since $\lim_{t\to\infty} \lambda_{xt,s} = 0$ and n^m_{1xt} is bounded, $\lim_{t\to\infty} \int \lambda_{xt,s} n^m_{1xt} dx = 0$.

$$\frac{\partial D_{s,t}(n^m_{1\cdot s})}{\partial n^m_{1xs}} = \lambda_{xs,t}, \tag{58}$$

$$\frac{\partial D_{s,t}(n^m_{1\cdot s})}{\partial s} = -\mathcal{D}_{s,t}(q_{a\cdot s}, q_{b\cdot s}, n^m_{1\cdot s}) - \int \lambda_{xs,t} f(x, q_{a\cdot s}, q_{b\cdot s}, n^m_{1\cdot s}, s) dx. \tag{59}$$

From (54), the value function has the following property

$$\mathcal{D}_{\tau,t}(q_{a\cdot\tau}, q_{b\cdot\tau}, n^m_{1\cdot\tau}) + \int \left[\lambda_{x\tau,t}(n^m_{1\cdot\tau}) f(x, q_{a\cdot\tau}, q_{b\cdot\tau}, n^m_{1\cdot\tau}, \tau) + n^m_{1x\tau}\dot{\lambda}_{x\tau,t} \right] dx = 0.$$

So,

$$D_{s,t}(n^m_{1\cdot s}) = \int \lambda_{xs,t} n^m_{1xs} dx. \tag{60}$$

Using transversality conditions, $\lim_{t\to\infty} D_t(n^m_{1\cdot t}) = \lim_{t\to\infty} \tilde{D}_t(n^m_{1\cdot t}) = 0$.

$$D_{t,t}(n^m_{1\cdot t}) = -\int_t^\infty \frac{dD_{s,t}(n^m_{1\cdot s})}{ds} ds, \tag{61}$$

$$\tilde{D}_{t,t}(n^m_{1\cdot t}) = -\int_t^\infty \frac{d\tilde{D}_{s,t}(\tilde{n}^m_{1\cdot s})}{ds} ds. \tag{62}$$

$$
\begin{aligned}
\frac{d\tilde{D}_{s,t}(\tilde{n}^m_{1\cdot s})}{ds} &= \frac{\partial\tilde{D}_{s,t}(\tilde{n}^m_{1\cdot s})}{\partial s} + \int \frac{\partial\tilde{D}_{s,t}(\tilde{n}^m_{1\cdot s})}{\partial n^m_{1xs}}\dot{\tilde{n}}^m_{1xs}dx \\
&= -\mathcal{D}_{s,t}(\tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}, \tilde{n}^m_{1\cdot s}) - \int \lambda_{xs,t}(\tilde{n}^m_{1\cdot s}) f(x, \tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}, \tilde{n}^m_{1\cdot s}, s)dx + \int \lambda_{xs,t}(\tilde{n}^m_{1\cdot s})\dot{\tilde{n}}^m_{1xs}dx \\
&= -\mathcal{D}_{s,t}(\tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}, \tilde{n}^m_{1\cdot s}) - \int \lambda_{xs,t}(\tilde{n}^m_{1\cdot s}) f(x, \tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}, \tilde{n}^m_{1\cdot s}, s)dx - \int \dot{\lambda}_{xs,t}(\tilde{n}^m_{1\cdot s})\tilde{n}^m_{1xs}dx \\
&+ \int \frac{d}{ds}\left[\lambda_{xs,t}(\tilde{n}^m_{1\cdot s})\tilde{n}^m_{1xs}\right]dx \\
&= -\mathcal{D}_{s,t}(\tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}, \tilde{n}^m_{1\cdot s}) - \int \lambda_{xs,t}(\tilde{n}^m_{1\cdot s}) f(x, \tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}, \tilde{n}^m_{1\cdot s}, s)dx + \int \frac{\partial\mathcal{H}_{s,t}}{\partial n^m_{1xs}}\tilde{n}^m_{1xs}dx \\
&+ \int \frac{d}{ds}\left[\lambda_{xs,t}(\tilde{n}^m_{1\cdot s})\tilde{n}^m_{1xs}\right]dx \\
&= -\mathcal{D}_{s,t}(\tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}, \tilde{n}^m_{1\cdot s}) - \int \lambda_{xs,t}(\tilde{n}^m_{1\cdot s}) f(x, \tilde{q}_{a\cdot s}, \tilde{q}_{b\cdot s}, \tilde{n}^m_{1\cdot s}, s)dx \\
&+ \mathcal{D}_{s,t}(q_{a\cdot s}, q_{b\cdot s}, n^m_{1\cdot s}) + \int \lambda_{xs,t}(n^m_{1\cdot s}) f(x, q_{a\cdot s}, q_{b\cdot s}, n^m_{1\cdot s}, s)dx \\
&+ \int \frac{d}{ds}\left[\lambda_{xs,t}(\tilde{n}^m_{1\cdot s})\tilde{n}^m_{1xs}\right]dx.
\end{aligned}
$$

The second equality follows from (58) and (58). The fourth equality uses the necessary condition (54). Because of (57), from the last equality,

$$
\frac{d\tilde{D}_{s,t}(\tilde{n}^m_{1\cdot s})}{ds} \geq \int \frac{d}{ds}\left[\lambda_{xs,t}(\tilde{n}^m_{1\cdot s})\tilde{n}^m_{1xs}\right]dx = \frac{dD_{s,t}(\tilde{n}^m_{1\cdot s})}{ds}. \tag{63}
$$

The equality follows from (60). Together with (61) and (62), (63) implies $\tilde{D}_{t,t}(n^m_{1\cdot t}) \leq D_{t,t}(n^m_{1\cdot t})$. Hence the sufficiency.

A.5 Proof of the Existence Theorem, Theorem 1

The proof consists of two parts. First, we will show that there exists an optimal stationary contract menu taking the market price as given. Second, we will show there exists a price for the competitive interdealer market that clears the market.

Denote the market price to be P. For any price P such that $P = d_\kappa$ for $\kappa \in (x_L, x_H)$, we will give the proof for the existence of a stationary ask contract menu. The proof for the bid contract menu follows symmetrically. Specifically, we will show that there exists a $(q_{a\cdot}, p_{a\cdot})$ such that $q_{ax} \in \varphi_x(q_{a\cdot})$ for all $x \in [\kappa, x_H]$, and p_{ax} follows equation (22) without the time subscript.

I use Kakutani-Fan-Glicksberg Corollary (P583, Section 17.55 of Aliprantis and Border (2006)[2]) to prove this claim. The corollary states that for a nonempty compact convex subset of a locally convex Hausdorff space, X, and a correspondence, $\varphi : X \rightrightarrows X$, that has closed graph and nonempty convex values, there exists a nonempty and compact set of fixed points of φ.

The set of functions, K, in this context is the set of increasing functions in the space $[\kappa, x_H]^{[0,q]}$, $K \equiv \{q_{a\cdot} : [\kappa, x_H] \to [0, q], q_{ax} \text{ is increasing in } x\}$. First, by Tychonoff Product Theorem (P52, 2.61 of Aliprantis and Border (2006)[2]), $[\kappa, x_H]^{[0,q]}$ is a compact set in the product topology. Since K is a closed and convex subset of the set, K is compact and convex in the product topology .

Clearly, the correspondence has nonempty convex values. What remains to show is that φ has closed graph. In other words, for a sequence of $\{q^n_{a\cdot}\}^{\infty}_{n=1}$ such that $q^n_{a\cdot} \in K$ and $\lim_{n\to\infty} q^n_{a\cdot} \in K$ in the product topology, we need to show that for any sequence $\{y^n_{\cdot}\}^{\infty}_{n=1}$such that $y^n_{\cdot} \in \varphi(q^n_{a\cdot})$, the limit exists in product topology and $y_{\cdot} \equiv \lim_{n\to\infty} y^n_{\cdot} \in K$.

To show this, first notice that virtual value $\hat{v}^n_{\cdot}$ that corresponds to $q^n_{a\cdot}$ converges to the virtual value $\hat{v}_{\cdot}$ that corresponds to $q_{a\cdot}$. From (29), $\hat{v}^n_x$ is a function of q^n_{ax} and integrals of q^n_{as} on $[\kappa, x]$ and $[x, x_H]$. Since $q^n_{a\cdot}$ converges pointwise to $q_{a\cdot}$ and $q^n_{a\cdot}$ is uniformly bounded, $q^n_{a\cdot}$ also weakly converges to $q_{a\cdot}$, which implies that the integrals of q^n_{as} on $[\kappa, x]$ and $[x, x_H]$ also converge to the integrals of q_{ax} on $[\kappa, x]$ and $[x, x_H]$. Therefore, $\hat{v}^n_{\cdot}$ converges pointwise to $\hat{v}_{\cdot}$. Using similar arguments, we can show $\hat{G}^n_{\cdot}$ converges pointwise to $\hat{G}_{\cdot}$. Therefore, $G^n_{\cdot}$ converges pointwise to $G_{\cdot}$.

Because v^n_x is continuous and weakly increasing in x, there exist closed intervals $[a^n, b^n], [a, b] \subseteq [\kappa, x_H]$, such that $v^n_x = 0$ if and only if $x \in [a^n, b^n]$ and $v_x = 0$ if and only if $x \in [a, b] \subseteq [\kappa, x_H]$. Because v^n_x converges pointwise to v_x, $a \leq \lim_{n\to\infty} a^n$ and $\lim_{n\to\infty} b^n \leq b$. Therefore, $\varphi(\lim_{n\to\infty} q^n_{a\cdot}) \subseteq \varphi(q_{a\cdot})$. Because $y^n_{\cdot} \in \varphi(q^n_{a\cdot})$ and $y_{\cdot} \in \varphi(\lim_{n\to\infty} q^n_{a\cdot})$, $y_{\cdot} \in \varphi(q_{a\cdot})$.

So far we showed that taking κ as given there exists a fixed point. Because we assume that $A = \frac{1}{2}$ and the distribution function $g(x)$ is symmetric, there exists an equilibrium trading frequency function $q_{bx} = q_{a,2x_M-x}$, which guarantees that the market clears at $\kappa = x_M$.

From (29), $\hat{v}_\kappa < 0$ and $\hat{v}_{x_H} > 0$. Because $\int_\kappa^x (r + \delta + \gamma + q_{au}) d'_u du$ and $\int_x^{x_H} n^m_{0u} du$ are continuous in x, $\hat{v}_x < 0$ for x close to κ and $\hat{v}_x > 0$ for x close to x_H. v_x inherits the property. So, the fixed point must satisfy the property that $q_{ax} = 0$ for x close to κ and $q_{ax} = q$ for x close to x_H. Likewise, $q_{bx} = 0$ for x close to κ and $q_{bx} = q$ for x close to x_L.

Using property 1 of Lemma 1, $q_{ax} = 0$ for all $x < \kappa$. Likewise, $q_{bx} = 0$ for all $x > \kappa$. This confirms Conjecture 1 and completes the proof. Q.E.D.

A.6 Proof of Corollary 1

From Theorem 1, we can show that $\bar{x}^b < \kappa < \underline{x}^a$. This means that dealers must earn positive profits. As long as dealers earn positive profits, there must exists non-trivial bid-ask spread. No trade takes place for traders with preference type $x \in \left[\bar{x}^b, \underline{x}^a\right]$. So, there are delays due to endogenous trading frictions and trading volume must decrease.

Under efficient allocation, there is no delay in trade for any trader. Denote κ^* to be the threshold type for the efficient allocation. If $\kappa^* \geq \kappa$, the flow of asset supply from sellers is smaller than the trading flow under complete information, because there exists trading delay for some seller types. Likewise, if $\kappa^* < \kappa$, the flow of asset demand from buyers is smaller than the trading flow under complete information. Therefore, the trading flow with asymmetric information must be lower than that with efficient allocation. Q.E.D.

A.7 Analytical Example 1 ($\gamma = 0$)

I present the solution for ask contracts here. The solution for bid contracts follows the same procedure. For $x \in (\underline{x}^a, \bar{x}^a)$, $q_{ax} \in (0, q)$. So, $\hat{v}_x = 0$, $\hat{v}_x$ satisfying (29). Suppose the solution to q_{ax} is strictly increasing in this region, a conjecture to be verified later, then, from (29), we have

$$\frac{\int_{\kappa}^{x} \left(r + \delta + q_{as}\right) d'_s ds}{\left(r + \delta + q_{ax}\right) d'_x} = \frac{\int_{x}^{x_H} \tilde{n}^m_{0s} ds}{\tilde{n}^m_{0x}}, \forall x \in (\underline{x}^a, \bar{x}^a), \tag{64}$$

where

$$\begin{aligned} d'_x &= \frac{1}{r + \delta + q_{ax}}, \\ \tilde{n}^m_{0x} &= g(x) \frac{\delta}{\delta + q_{ax}} (1 - A). \end{aligned}$$

Let $F(x) = \int_x^{x_H} \frac{g(s)}{\delta + q_{as}} ds$. The ordinary differential equation (64) can be simplified to be

$$\begin{aligned} \frac{d}{dx} \ln F(x) &= -\frac{1}{x - \kappa}, \forall x \in (\underline{x}^a, \bar{x}^a), \\ F(\bar{x}^a) &= \frac{1 - G(\bar{x}^a)}{\delta + q}. \end{aligned}$$

The boundary condition at $x = \bar{x}^a$ uses the fact that $q_{ax} = q$, for all $x > \bar{x}^a$. From the simplified ODE, we can solve for q_{ax}. $\forall x \in (\underline{x}^a, \bar{x}^a)$,

$$q_{ax} = g(x) \frac{(x - \kappa)^2}{\bar{x}_a - \kappa} \frac{\delta + q}{1 - G(\bar{x}_a)} - \delta. \tag{65}$$

At $\bar{x}_a$, $q_{a\bar{x}^a} = q$. So, from equation (65), $\bar{x}_a$ satisfies the following equation: $1 = \frac{g(\bar{x}^a)(\bar{x}^a-\kappa)}{1-G(\bar{x}^a)}$. At $\underline{x}^a$, $q_{a\underline{x}^a} = 0$. From equation (65), $\underline{x}_a$ satisfies the following equation: $\frac{\delta}{\delta+q} = \left(\frac{\underline{x}^a-\kappa}{\bar{x}^a-\kappa}\right)^2 \frac{g(\underline{x}^a)}{g(\bar{x}^a)}$. The spread $p_{ax} - P$ satisfies the following equation,

$$
\begin{aligned}
q_{ax}(p_{ax} - P) &= \int_{\kappa}^{x} (q_{ax} - q_{as}) d_s^{'} ds \\
&= \int_{\underline{x}^a}^{x} \frac{g(x)\,(x-\kappa)^2 - g(s)(s-\kappa)^2}{g(\bar{x}^a)\,(\bar{x}^a-\kappa)^2} \frac{\delta+q}{r + \frac{g(s)}{g(\bar{x}^a)}\left(\frac{s-\kappa}{\bar{x}^a-\kappa}\right)^2(\delta+q)} ds + q_{ax}\frac{(\underline{x}^a-\kappa)}{r+\delta},
\end{aligned}
$$

which implies that

$$
p_{ax} - P = \frac{\underline{x}^a-\kappa}{r+\delta} + \frac{1}{q_{ax}} \int_{\underline{x}^a}^{x} \frac{g(x)\,(x-\kappa)^2 - g(s)(s-\kappa)^2}{g(\bar{x}^a)\,(\bar{x}^a-\kappa)^2} \frac{\delta+q}{r + \frac{g(s)}{g(\bar{x}^a)}\left(\frac{s-\kappa}{\bar{x}^a-\kappa}\right)^2(\delta+q)} ds.
$$

If $r \ll 1$,

$$
p_{ax} - P \;=\; \frac{\underline{x}^a-\kappa}{\delta} + \frac{1}{q_{ax}} \int_{\underline{x}^a}^{x} \left[\frac{g(x)\,(x-\kappa)^2}{g(s)(s-\kappa)^2} - 1\right] ds.
$$

A.8 Bunching equilibrium in Example 1

Assume $G \in \Delta[0,1]$ and $G(x) = x$.

Suppose there exits $x_a \in (\kappa, 1)$ such that $q_{ax} = 0$ for $x \in [\kappa, x_a)$ and $q_{ax} = q$ for $x \in (x_a, 1]$. Then, for $x < x_a$, $d'_x = \frac{1}{r+\delta}$, $n^m_{0x} = 1 - A$. For $x > x_a$, $d'_x = \frac{1}{r+\delta+q}$, $n^m_{0x} = \frac{\delta}{\delta+q}(1-A)$. From equation 29, for $x < x_a$,

$$
\begin{aligned}
\hat{v}_x &= (x - \frac{1}{2})\frac{1}{r+\delta} - \frac{1}{r+\delta}\left(x_a - x + (1-x_a)\frac{\delta}{\delta+q}\right) \\
&= \frac{2}{r+\delta}x - \frac{1}{2(r+\delta)} - \frac{1}{r+\delta}\frac{\delta+qx_a}{\delta+q} \\
&= \frac{1}{r+\delta}\left(2x - \frac{\delta+qx_a}{\delta+q} - \frac{1}{2}\right)
\end{aligned}
$$

and for $x > x_a$,

$$
\begin{aligned}
\hat{v}_x &= \frac{x-x_a}{r+\delta+q} + \frac{r+\delta}{r+\delta+q}\frac{x_a-\frac{1}{2}}{r+\delta} - \frac{1}{r+\delta+q}(1-x) \\
&= \frac{1}{r+\delta+q}\left(2x - \frac{3}{2}\right)
\end{aligned}
$$

Because $\hat{v}_x$ is strictly increasing on intervals $[\kappa, x_a)$ and $(x_a, x_H]$, $\hat{G}(x) = \int_\kappa^x \hat{v}_x dx$ is convex on $[\kappa, x_a)$ and $(x_a, x_H]$, but $\hat{G}(x)$ is not convex in the neighborhood of x_a.

Then, there must exist a, b, such that $\kappa \leq a < x_a < b \leq x_H$, and

$$\hat{v}_x = v, \text{ for } x = a, b$$
$$\int_a^b \hat{v}_x dx = v(b-a)$$

Because $\varphi(x_a) = [0, q]$, it must be that $v = 0$. So, a, b, x_a are solved jointly by the following equations

$$\hat{v}_x = 0, \text{ for } x = a, b \tag{66}$$
$$\int_a^b \hat{v}_x dx = 0. \tag{67}$$

$\hat{v}_a = 0$ iff $a = \frac{1}{4} + \frac{1}{2}\frac{\delta + q x_a}{\delta + q}$. We can check that $a < x_a$ if and only if $\frac{3\delta+q}{2(2\delta+q)} \leq x_a$. $\hat{v}_b = 0$ iff $b = \frac{3}{4}$. And $x_a < \frac{3}{4}$. We can check that $\frac{3\delta+q}{2(2\delta+q)} < \frac{3}{4}$. From equation (67)

$$\frac{1}{r+\delta}\left[x_a - \left(\frac{1}{4} + \frac{1}{2}\frac{\delta + q x_a}{\delta + q}\right)\right]^2 = \frac{1}{r+\delta+q}\left(\frac{3}{4} - x_a\right)^2.$$

$$x_a = \left[1 + \sqrt{\frac{r+\delta}{r+\delta+q}}\frac{2\delta+q}{2(\delta+q)}\right]^{-1}\left[\frac{3}{4} + \sqrt{\frac{r+\delta}{r+\delta+q}}\left(\frac{1}{4} + \frac{1}{2}\frac{\delta}{\delta+q}\right)\right].$$

A.9 Analytical Example 2 (Discrete Distribution)

Assume that the preference type follows a discrete distribution on $x_L < y < x_M < z < x_H$. $\Pr(x = x_L) = \Pr(x = x_H) = \pi_1$, $\Pr(x = y) = \Pr(x = z) = \pi_2$. $A = \frac{1}{2}$. $x_M = \frac{1}{2}(x_L + x_H)$, $x_M - y = z - x_M$. Because the distribution is symmetric, I will focus on solving the ask contracts. Because of symmetry, $\kappa = x_M = P$. Dealers earn no profit selling assets to traders of type x_M. So, $q_{a\kappa t} = 0$. And there is no point to delay the trade with traders of type x_H. So, $q_{ax_Ht} = q$. Because type z traders are the lowest type dealers are going to sell assets to, $p_{azt} = d_{zt}$. And since it is optimal for dealers to make the IC constraint of type x_H traders bind, $q(d_{xHt} - p_{xHt}) = q_{azt}(d_{xHt} - d_{zt})$. The first order condition of the problem at the steady state is

$$(d_z - P)n_{0z}^m \frac{r+\delta+\gamma}{r+\delta+\gamma+q_{az}} - (d_{xH} - d_z)n_{0xH}^m - \gamma_{1az} + \gamma_{0az} = 0,$$

where $n_{0z}^m = \pi_2 \frac{\alpha}{\alpha+\gamma}\frac{\alpha+\delta+\gamma}{\alpha+\delta+\gamma+\frac{\alpha+\delta}{\delta}q_{az}}(1-A)$. Assume that $r \ll \delta$.

$$q_{az} = \left[\frac{z - x_M}{x_H - z}\frac{\pi_2}{\pi_1}\frac{(\alpha+\delta)\left(1+\frac{q}{\delta}\right)+\gamma}{\alpha+\delta+\gamma} - 1\right](\delta+\gamma).$$

A.10 Proof of Corollary 3

The proof is derived from equation (34). When $\alpha \gg 1$, $q_{az} = \left(\frac{z-x_M}{x_H-z}\frac{\pi_2}{\pi_1} - 1\right)(\delta+\gamma)$. As in Example 2, I assume that $q_{az} \in (0, q)$, which implies that $\frac{z-x_M}{x_H-z}\frac{\pi_2}{\pi_1} > 1$. Therefore, q_{az} is increasing in γ. When $\alpha \ll \delta$, $q_{az} = \left(\frac{z-x_M}{x_H-z}\frac{\pi_2}{\pi_1} - 1\right)(\delta+\gamma) + \frac{z-x_M}{x_H-z}\frac{\pi_2}{\pi_1}q$. If $q_{az} \in (0, q)$, it must be that $\frac{z-x_M}{x_H-z}\frac{\pi_2}{\pi_1} < 1$. Therefore, q_{az} is decreasing in γ. Q.E.D.

A.11 Proof for Corollary 4

For δ_1, δ_2 such that $\delta_2 > \delta_1 > 0$, denote the trading frequency and cutoffs of δ_j by q_{axj}, $\bar{x}^a_j$ and $\underline{x}^a_j$. From equation (32) and (33), $\bar{x}^a_1 = \bar{x}^a_2$, $\underline{x}^a_1 < \underline{x}^a_2$. For all $x \in (\underline{x}^a_2, \bar{x}^a_2)$, $\frac{d}{dx}(q_{ax2} - q_{ax1}) = \frac{g(x)}{g(\bar{x}^a)}\left(\frac{x-x_M}{\bar{x}^a-x_M}\right)^2(\delta_2-\delta_1) > 0$. Since $q_{ax2} - q_{ax1} < 0$ at $\underline{x}^a_2$ and $q_{ax2} - q_{ax1} = 0$ at $x = \bar{x}^a_2$, $q_{ax2} < q_{ax1}$ for all $x \in (\underline{x}^a_2, \bar{x}^a_2)$. Q.E.D.

A.12 Proof for Corollary 5

From equation (34), $q_{az} = (\phi-1)(\delta+\gamma) + q\frac{\varphi}{1-\frac{\alpha\gamma}{(\alpha+\delta)(\gamma+\delta)}}$. For q_{az} to be between 0 and q, $\phi < 1$. Therefore, q_{az} is decreasing in δ. Q.E.D.

A.13 Proof for Corollary 6

The Corollary holds for Example 2, according to equation (34). For Example 1, consider ζ_1 and ζ_2 such that $\zeta_2 > \zeta_1 > 1$. Denote the trading frequency and cutoffs corresponding to ζ_j by q_{axj}, $\bar{x}^a_j$ and $\underline{x}^a_j$.

According to equation (32), $\bar{x}^a_2 > \bar{x}^a_1$. According to equation (33), $\underline{x}^a_2 > \underline{x}^a_1$ because

$$\underline{x}^a_2 - x_M = \left(\frac{\delta}{\delta+q}\right)^{\frac{1}{2+\zeta_2}}(\bar{x}^a_2 - x_M) > \left(\frac{\delta}{\delta+q}\right)^{\frac{1}{2+\zeta_1}}(\bar{x}^a_1 - x_M) = \underline{x}^a_1 - x_M.$$

For any $x \in [\underline{x}^a_2, \bar{x}^a_1]$, according to equation (31),

$$q_{ax2} = \left(\frac{x-x_M}{\bar{x}^a-x_M}\right)^{2+\zeta_2}(\delta+q) - \delta < \left(\frac{x-x_M}{\bar{x}^a-x_M}\right)^{2+\zeta_1}(\delta+q) - \delta = q_{ax1}.$$

Q.E.D.

A.14 Proof of Lemma 5

We focus without loss of generality on the ask contracts. By Lemma 4

$$F_a(R_{0x}) = q_{ax}/q,\ f_a(R_{0x})R'_{0x} = q'_{ax}/q,$$

and

$$\int_{R_{0\kappa}}^{R_{0x}} RdF_a(R) = p_{ax}q_{ax}/q,$$
$$q\int_{R_{0\kappa}}^{R_{0x}} RdF_a(R) = (r+\delta+q_{ax}+\gamma)\left[P+\int_\kappa^x \frac{1+\gamma e'_s}{r+\delta+q_{as}+\gamma}ds\right] - (x+\delta Ed+\gamma e_x),$$
$$qR_{0x}f_a(x)R'_{0x} = q'_{ax}\left[P+\int_\kappa^x \frac{1+\gamma e'_s}{r+\delta+q_{as}+\gamma}ds\right].$$

Since $R'_{0x}f_0(R_{0x}) = q'_{ax}/q$, we have, $R_{0x} = P + \int_\kappa^x \frac{1+\gamma e'_s}{r+\delta+q_{as}+\gamma}ds$. Since $d'_x = \frac{1+\gamma\frac{1}{r+\alpha+\delta}}{r+\delta+q_{bx}+q_{ax}+\gamma\frac{r+\delta}{r+\alpha+\delta}}$ and $P = d_\kappa$, $P + \int_\kappa^x \frac{1+\gamma e'_s}{r+\delta+q_{as}+\gamma}ds = d_\kappa + \int_\kappa^x d'_s ds = d_x$. Therefore, $R_{0x} = d_x = M_{1x} - M_{0x}$.Q.E.D.

References

[1] Gara Afonso, Anna Kovner, and Antoinette Schoar. Trading partners in the interbank lending market. 2014.

[2] Charalambos D Aliprantis and Kim Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Science & Business Media, 2006.

[3] Max R Blouin and Roberto Serrano. A decentralized market with common values uncertainty: Non-steady states. *Review of Economic Studies*, pages 323–346, 2001.

[4] Kenneth Burdett and Kenneth L Judd. Equilibrium price dispersion. *Econometrica*, pages 955–969, 1983.

[5] Braz Camargo and Benjamin Lester. Trading dynamics in decentralized markets with adverse selection. *Journal of Economic Theory*, 153:534–568, 2014.

[6] Briana Chang. Adverse selection and liquidity distortion in decentralized markets. *working paper*, 2011.

[7] Jonathan Chiu and Thorsten Koeppl. Trading dynamics with adverse selection and search: Market freeze, intervention and recovery. 2011.

[8] Marco Di Maggio, Amir Kermani, and Zhaogang Song. The value of trading relationships in turbulent times. Technical report, National Bureau of Economic Research, 2016.

[9] Darrell Duffie, Nicolae Gârleanu, and Lasse Heje Pedersen. Over-the-counter markets. *Econometrica*, 73(6):1815–1847, 2005.

[10] Darrell Duffie, Semyon Malamud, and Gustavo Manso. Information percolation in segmented markets. *Journal of Economic Theory*, 153:1–32, 2014.

[11] Amy K Edwards, Lawrence E Harris, and Michael S Piwowar. Corporate bond market transaction costs and transparency. *The Journal of Finance*, 62(3):1421–1451, 2007.

[12] William Fuchs and Andrzej Skrzypacz. Bargaining with arrival of new traders. *The American Economic Review*, pages 802–836, 2010.

[13] Daniel Garrett. Incoming demand with private uncertainty. *Unpublished manuscript, Toulouse School of Economics*, 2013.

[14] Mikhail Golosov, Guido Lorenzoni, and Aleh Tsyvinski. Decentralized trading with private information. *Econometrica*, 82(3):1055–1091, 2014.

[15] Richard C Green. Presidential address: Issuers, underwriter syndicates, and aftermarket transparency. *The Journal of Finance*, 62(4):1529–1550, 2007.

[16] Richard C Green, Burton Hollifield, and Norman Schürhoff. Dealer intermediation and price behavior in the aftermarket for new bond issues. *Journal of Financial Economics*, 86(3):643–682, 2007.

[17] Veronica Guerrieri and Robert Shimer. Dynamic adverse selection: A theory of illiquidity, fire sales, and flight to quality. *The American Economic Review*, 104(7):1875–1908, 2014.

[18] Veronica Guerrieri, Robert Shimer, Randall Wright, et al. Adverse selection in competitive search equilibrium. *Econometrica*, 78(6):1823–1862, 2010.

[19] Faruk Gul, Hugo Sonnenschein, and Robert Wilson. Foundations of dynamic monopoly and the coase conjecture. *Journal of Economic Theory*, 39(1):155–190, 1986.

[20] Igal Hendel and Alessandro Lizzeri. Adverse selection in durable goods markets. *American economic review*, 89(5):1097–1115, 1999.

[21] Johannes Hörner and Nicolas Vieille. Public vs. private offers in the market for lemons. *Econometrica*, 77(1):29–69, 2009.

[22] Julien Hugonnier, Benjamin Lester, and Pierre-Olivier Weill. Heterogeneity in decentralized asset markets. Technical report, National Bureau of Economic Research, 2014.

[23] Roman Inderst. Matching markets with adverse selection. *Journal of Economic Theory*, 121(2):145–166, 2005.

[24] Bernd Kirchheim and Jan Kristensen. Differentiability of convex envelopes. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 333(8):725–728, 2001.

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

[26] Ricardo Lagos and Shengxing Zhang. Monetary exchange in over-the-counter markets: A theory of speculative bubbles, the fed model, and self-fulfilling liquidity crises. Technical report, National Bureau of Economic Research, 2015.

[27] Stephan Lauermann and Asher Wolinsky. Search with adverse selection. *working paper*, 2011.

[28] Benjamin Lester, Guillaume Rocheteau, and Pierre-Olivier Weill. Competing for order flow in otc markets. *Journal of Money, Credit and Banking*, 47(S2):77–126, 2015.

[29] Dan Li and Norman Schürhoff. Dealer networks. 2014.

[30] Daniel Liberzon. *Calculus of variations and optimal control theory: a concise introduction.* Princeton University Press, 2012.

[31] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

[32] E. Pagnotta and T. Philippon. Competing on speed. 2012.

[33] Christopher A Pissarides. *Equilibrium unemployment theory.* MIT press, 1990.

[34] Asher Wolinsky. Information revelation in a market with pairwise meetings. *Econometrica*, pages 1–23, 1990.