Endogenous Market-making and Formation of Trading Links

Briana Chang[*] Shengxing Zhang[†]

University of Wisconsin–Madison London School of Economics

February 15, 2015

Abstract

This paper develops a dynamic matching model to analyze the equilibrium trading structure in OTC markets. All traders have the same trading technology, and they optimally choose whom to connect to as well as whether to remain active in each period. We show that traders with more volatile preference (i.e. with higher needs for trade) always choose to match with traders with more stable preference (i.e., with less needs for trade) and they also leave the market earlier. The model therefore endogenously generates a core-periphery market structure, where traders with less trading needs develop more trading links in equilibrium and act like market makers. As the role of market-making is endogenous, we therefore provide an answer to why customers choose to trade with dealers instead of trading among themselves, and we can analyze how the market-making profit depends on underlying frictions.

Keyword: over-the-counter market, core-periphery trading network, matching, endogenous intermediation

JEL classification: C70, G12, G21

[*]Email: bchang@bus.wisc.edu.

[†]Email: S.Zhang31@lse.ac.uk. Zhang thanks the support from the Centre for Macroeconomics at LSE.

1 Introduction

Trading in decentralized markets requires traders to find the counterparty. In reality, we observe that traders often choose to trade with dealers instead of looking for the better counterparty themselves. For example, Li and Schurhoff (2011)[17] show that dealers intermediate 94% of the trades in the municipal bond market, with most of the intermediated trades representing customer-dealer-customer transactions. What remains unknown is why such trading structure arises in the first place. In particular, why traders who have higher needs for trade ("customers") choose to trade with traders who have lower needs for trade ("dealers")?

To answer this question properly, we develop a dynamic matching model which has two distinct properties from the existing literature: First, all traders have the same trading technology and capacity. This is the key difference from most micro-structure literature, where the "market-makers" are often assumed to have certain advantage with their trading technology. Second, traders optimally choose whom to contact as well as whether to be active in the market in each period. In other words, our dynamic model generates an endogenous trading network, where all trading links as well as the number of links for each trader are determined in equilibrium. With these features, we can therefore explicitly answer why "customers" choose to trader with "dealers" instead of trading among themselves. Furthermore, the model sheds light on "who" will be at the core of a trading network.

The key heterogeneity in our model is traders' exposure to preference shocks. Traders who are subject to more volatile shocks have more extreme preference realization, and therefore are the ones who have higher needs for trade. One can interpret these traders as the ones who have higher risk-sharing needs. On the other hand, traders with more stable positions are the ones who gain less from participating in the market. For example, in the interbank lending market, the volatility represents liquidity shortage that arises from deposit withdrawal, or large payment from one bank to another. Some banks may have more volatile liquidity need than others. This could be because they are small in size or their customer basis is less stable. On the other hand, we know that dealer banks are typically large banks, who themselves do not need to borrow or lend from other banks.

The main decision in our model is the formation of trading links. Before the realization of preference shocks, traders choose whom to match with each period, and agree on the allocation and the transfer contingent on the possible realization within each match. We also allow them to choose to match with no one, which can then be interpreted as if they leave the market. This formation decision thus gives the underline trading structure. For any realization, traders simply trade based on the trading rule they agree on.

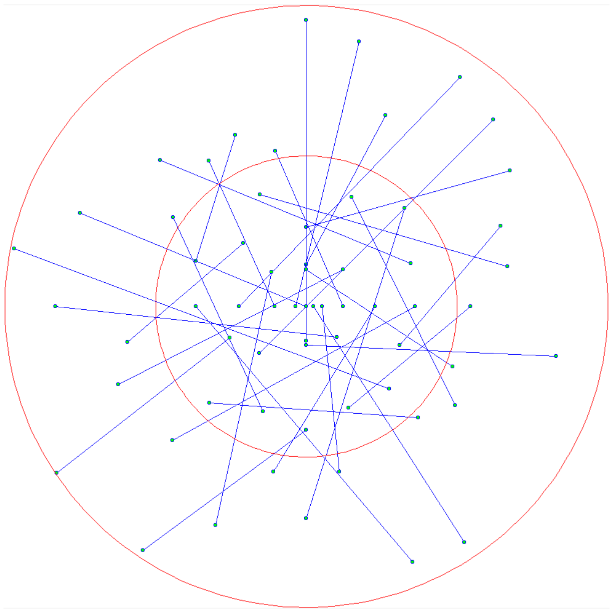
We show that, in equilibrium, traders with more volatile preference always match with traders with less volatile preference, this is true even though traders are risk-neutral and preferences are negatively correlated. For the sake of illustration, consider a simple one-period example with two types: the preference realization of the extreme types is either -1 or 1 with equal probability, while the realization of the stable types is always zero. Note that, as long as the preference shocks among extreme guys are weekly negatively correlated, an extreme type is a better counterparty for another extreme type in the sense that they can generate a higher pairwise surplus. However, we show that the only stable matching outcome is pairing the extreme type with the stable type. Such result can be extended to an environment with continuous types, dynamic trading, and heterogeneous correlation.

The intuition is very simple: matching involving the stable type provides insurance against the state where extreme types have similar realizations, where the gain from trade between extreme types is small. That is, the stable trader acts as a "market-maker", as they are always ready to take the opposite position of the extreme traders. Although the market maker himself might not need to trade, and despite that the "customers" can reach a higher pairwise surplus with another customer, trading through market makers minimizes the uncertainty of the preference shocks in the economy and therefore maximizes the aggregate surplus. As it is well-known that the matching outcome must have the core property, the equilibrium surplus sharing rule therefore guarantees that it is indeed optimal for extreme types to match with stable types in equilibrium.

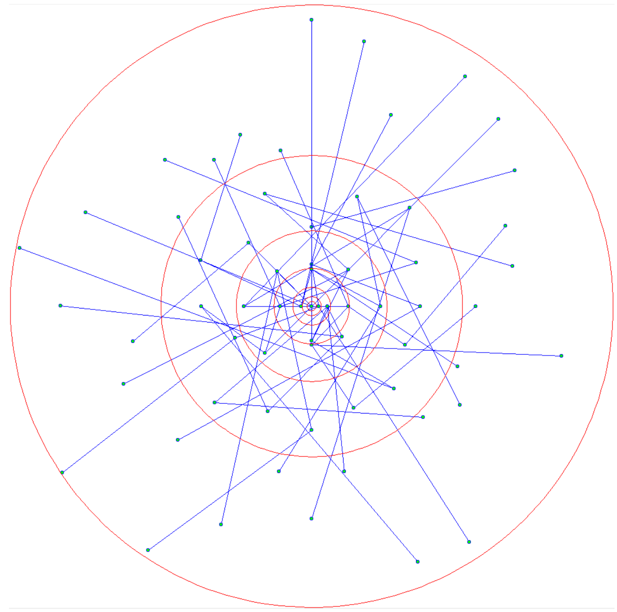
This insight carries through in a dynamic model with N-periods. N can be interpreted as potential trading opportunities in the market, or liquidity of the market. As in the static model, a trader with more volatile preference always matches with a more stable type. Moreover, it is optimal for him to leave the market afterward, as he has already reaches his first-best allocation, given the matching rule. The matching outcome for the next period thus features the same trading pattern with an updated distribution of active traders. This recursive property makes our dynamic model tractable. We can therefore analytically characterize traders' equilibrium payoff, matching decisions, as well as trading dynamics over time for any N.

Figure 1 illustrates the matching outcome. A point in the interior of the circle represents a trader and his type is given by the distance from the center to the point.Hence, the further away from the center, the more volatile the trader's preference. The line simple represents the link between two traders. Figure 1(a) shows the result for one-period model: there exists a cutoff type such that traders above the cutoff type (act like "customers") randomly match with traders below the cutoff type ("market-maker").

In a dynamic model with multiple rounds, we call a trader "dealer" if, given the realization, he trades on both directions (i.e., both buy and sell) on the equilibrium path. Figure 1(b) shows the outcome for the dynamic model: there is a time-varying cutoff that divides "customers" (relatively volatile types) and "dealers" (relatively stable types) in each period. In equilibrium, "customers" always leave the market after he matches with "dealers". That explains why traders that are far away from the center have less links, and vice versa. Hence, traders who have more stable preference stay in the market longer and have more trading links in equilibrium. More importantly, since they always match with a more volatile type in each period, they simply trade based on customers' needs.



(a) Network in an illiquid market. ($N = 1$)



(b) Network in a liquid market. ($N = 7$)

Figure 1: Liquidity and network structure.

The model therefore endogenously generates a core-periphery structure. Traders who do not need to trade turns out to be the core of the network: they are most connected and have the highest gross trade volume. Moreover, this core-periphery structure implies heterogeneity in dealers' connectedness. Most volatiles types are purely customers, as they only trade once and leave the market right away. Traders with mid-range volatility act like "peripheral" dealers in the sense that they serve customer in the first period, but then they match with a "central" dealer next period. The model also implies that "central" dealers gained highest intermediation profits over time. All of these predictions are consistent with the hierarchical structure documented by Li and Schürhoff (2014)[17] and Neklyudov 2013 ([18]).[1]

Since the role of market-making is endogenous, we can then analyze how the intermediation profit (which maps to bid-ask spread) changes as a function of the underlying friction. We consider two key frictions: potential trading opportunities and correlation of traders' preferences. Trading opportunities captures the fact that trading process or building connections can be costly. On the other hand, since we allow for traders to choose their trading partner based on the volatility, the correlation of traders' preference then simply

[1]Li and Schürhoff (2014)[17] found that, in municipal bond markets, there is a higher level of heterogeneity among dealers in terms of connectedness and trading costs increase strongly with dealer centrality. Neklyudov (2013)[18] found that more efficient central dealers provide intermediation services to peripheral dealers.

captures the information friction. When preferences are perfectly negatively correlated, traders know for sure who is their best counterparty. Infinite trading opportunities and perfectly negative correlations thus represent frictionless markets. We show that the profit of market making indeed increases with underlying friction. In the limit case of a frictionless market, traders' payoff converges to the payoff implied by a competitive centralized market. That is, zero bid-ask spread.

These two frictions are in fact implicit assumptions in the standard random search and matching model, where the "right" counterparty arrives only at an exogenous rate (see Duffie, Garleanu and Pederson (2005)[9], Afonso and Lagos (2014)[2], Hugonnier, Lestor and Weill (2014)[13]). The key difference here is that traders choose whom to match as well as how many to match. We therefore have two distinct predictions: trades are concentrated among certain traders and this "matching" rate is endogenous and is heterogeneous across traders (see Rubinstein and Wollinsky (1987)[20]). Our paper is also related to Atkeson Eisfeldt and Weill (2014)[4] , where banks are different in terms of their size and marginal value of assets and all banks match with all other banks. In a static model, they show that large banks endogenously become dealers in the sense that large banks have highest gross notional of trade. Although we do not explicitly model banks size, one can interpret large banks have a more diversified profile and therefore have less exposure in their preference shocks. Our model has similar implications in the sense that trading volume is concentrated in those banks. Nevertheless, the economic here is very different, as we allow traders to choose whom to match, and we show that customers optimally direct their trade to dealers.

In contrast to models with exogenous network structure, where certain traders are assumed to be more connected and therefore "customers" must trade through "dealers", all "customers" in our model potentially can directly trade with another "customer". Nevertheless, it is indeed optimal for customers to build links to dealers and become less active in the market. Since our dynamic matching model endogenises the trading structure, our paper is also related to recent literature on the network formation. (Babus (2012)[6], Gale and Kariv(2007)[11], Faboodi (2014)[10].)[2] There are few papers on this area, where the formation is driven by different considerations: For example, Babus (2012)[6] allows trader to build relationship to overcome information frictions. Farboodi (2014)[10] consider two types of banks: banks who make risky investments over-connect, while banks who mainly provide funding end up with too few connections. We are the first paper to show that why stable types help create aggregate trading surplus and thus become the core of the network. Furthermore, despite the network structure, our model admits an analytical solution with rich heterogeneity, which facilitates characterizing asset price, trading volume, and trading dynamic.

The rest of the paper is organised as follows. We start from a static model to explain the main mechanism behind sorting on volatility and endogenous market-making in Section 2. We then extend the basic model to a dynamic setting in Section 3, where endogenous trading links generate a core-periphery trading structure.

[2]See Jackson (2005)[14] for detailed review on the literature on network formation.

Section 4 further studies implications of the dynamic model on properties of the trading network in OTC markets and market liquidity. Section 5 concludes the paper.

2 Basic model – endogenous market-making

2.1 Setup

The environment: There are two periods ($t = 0, 1$).There is an atomless continuum of risk-netural traders. Agents are indexed by their preference volatility $\sigma \in \Sigma = [\sigma_L, \sigma_H]$, which is exogenously given and publicly observable. And $G(\sigma)$ denotes the measure of traders with preference volatility below σ. There is one divisible asset, and all agents are endowed with a units of this asset at $t = 0$. The preference over the asset is realized at $t = 1$, is $u(\varepsilon, \tilde{a}) = \varepsilon\tilde{a}\mathbb{I}(0 \leq \tilde{a} \leq 2a)$, where ε is the realized preference type of the trader, and $\tilde{a}$ is the asset holding of the trade at period 1. The marginal return from holding an asset is zero when the asset holding of the trader exceeds $2a$. Let ε^i_σ denote the realized preference type for trader i with volatility σ, which is given by

$$\varepsilon^i_\sigma = c + x^i\sigma,$$

where $c \geq \sigma_H$ and x^i is a random variable that takes the value $x^i = \{-1, 1\}$ with equal probability at $t = 1$. For the sake of illustration, we start with the simple case that there is no correlation among traders so that $p \equiv \Pr(x^i x^j < 0) = \frac{1}{2}$ for $\forall\ j \neq i$. This is then the special case that each trader receives i.i.d preference shock. Our result remain intact for any parameter $p \in [0, 1)$. In Section x, we further impose more structure on traders' preference to formalize the general environment, where correlation can be anything between 1 and -1, and it can also be heterogenous across traders. Nevertheless, as all these extensions can be nested in our basic setup by setting different p, we derive our result for any given parameter p below. The volatility type and the asset holding of a trader are observable to other traders, but not his realised preference type.

Action Set: At $t = 0$, each trader can choose to match with another trader, and they agree on the contract that specifies the asset allocation and transfers contingent on the realization of shocks at $t = 1$. Hence, the object of interest here is the allocation as well as the optimal trading partner in equilibrium.

Our assumptions on preference and the information structure capture two things. First, the more volatile types have higher gains from trade. One can interpret those are the traders who have higher risk-sharing needs. On the other hand, the relatively stable types have lower exposure to such risks. One can think of them are, for example, the larger banks who have more diversified portfolio and therefore have less needs to engage risk-sharing in asset markets. Second, the fact that volatility is observable but the preference is uncertain –as long as preference are not perfectly correlated correlation–captures the idea that traders know who have higher risk-sharing needs, but they do not know for sure who will take the opposite position.

The assumptions on preference and the information structure are motivated by empirical works on the trading partners in OTC markets. In Afonso, Kovner and Schoar (2013)[3] on trading partners in the interbank lending market ([3]), they show that banks with trading relationships borrow and lend to each other and the fluctuations of their liquidity needs tend to be negatively correlated.[3] Our model is able to capture these empirical facts with these assumptions. The idea that traders might not know who is the best couterparty is also the key implicit assumption in all random search models (see [9, 15, 2, 13]). The key difference here is that we allow traders to choose their trading partners optimally, as they can act upon the information given by observable characteristics of their counterparties, such as preference volatility, asset holding and correlation with their own preference shocks.

2.2 Equilibrium

To facilitate the equilibrium definition, we now introduce notations for the contract and the payoff. Let set $S_{zz'}$ represent the possible realizations of preference within pair $(z, z') = (\sigma, a, \sigma', a')$. The contracts exchanged within pair (z, z') is then denoted by $\tau = \{(a(z, s), t(z, s)), a(z', s), t(z', s))\}$, which specified the amount of asset $a(z, s)$ and payment $t(z, s)$ received by agent σ in state $s = \{\varepsilon_\sigma, \varepsilon_{\sigma'}\} \in S_{\sigma\sigma'}$. Define Γ be the set of feasible contracts, $\tau \in \Gamma$ if and only if, for every s,

$$\begin{aligned} t(z, s) + t(z', s) &= 0 \\ a(z, s) + a(z', s) &\leq a + a'. \end{aligned}$$

For any $\tau \in \Gamma$, let $W(z|z')$ denote the expected value for trader$z$ who matches with z', which yields

$$W(z|z') = E_s \left[\varepsilon_\sigma a(z, s) + t(z, s)\right].$$

The maximum surplus generated within the pair (z, z'), denoted by $\Omega(z, z')$, is then given by

$$\Omega(z, z') = \max_{\tau \in \bar{\Gamma}} W(z|z') + W(z'|z)$$

As what matters is how the gain $\Omega(\sigma, \sigma')$ is divided among the pair, we work on traders' payoff directly instead of specifying the details of the contract. Let $f(\sigma, \sigma')$ denote the measure of the pair (σ, σ'). Hence, if $f(\sigma, \sigma') = 0$, we say that agents σ and σ' are not paired.

Definition 1 *An equilibrium is an allocation function $f : \mathbb{Z} \times \mathbb{Z} \to R_+$ and equilibrium payoff $W^*(\cdot) : \mathbb{Z} \to R_+$ satisfying the following conditions:*

[3]See table 3 of the paper ([3]), "determinants of existing relationships". Two key variables are among the determinants of trading relationships. (1) "symmetry", which measures whether banks in a relationship both borrow and lend to each other. The sign of the variable is significantly positive. This implies that the preference of banks in a relationship fluctuates. (2) "correlation of net customer funds", which measures whether the correlation of the volatility type for banks in a relationship. The sign of this variable is significantly negative, which means that volatility types of banks are negatively correlated.

1) Optimality for Traders:

$$W^*(z) \geq \max_{\tilde{z} \in \mathbb{Z}} \Omega(z, \tilde{z}) - W^*(\tilde{z}) \tag{1}$$

and for any $f(z, z') > 0$, $z' \in \arg\max_{z \in \mathbb{Z}} \{\Omega(z, z') - W^*(z)\}$.

2) Feasibility constraint:

$$\int f(z, \tilde{z}) d\tilde{z} = h(z) \text{ for } \forall z,$$

where $h(z)$ is the density function of z.

$\mathbb{Z}$ in the definition represents set of observable characteristics of an trader. In the basic model, to highlight the role of volatility of the trader's preference type, we focus on the case where every trader has the same asset holding. So in the baseline model, $\mathbb{Z} = \Sigma \cup \{\emptyset\}$ and a trader's type is the volatility of his preference, σ. $\{\emptyset\}$ is meant to allow traders to be matched with nobody. Or to say, traders are allowed not to participate in the matching game. And in the basic model, $h(z) = g(\sigma)$.

More generally, the sorting can be based on more observable characteristics. For example, when traders have heterogeneous asset holdings, the observable characteristics include the asset holding and the volatility type of an trader. In this case, $\mathbb{Z} = \Sigma \cup \{\emptyset\} \times [0, 2a]$. The type space also capture the correlation of volatility across groups of traders. In that case, $\mathbb{Z} = \Sigma \cup \{\emptyset\} \times [0, 2a] \times \mathbb{T}$, where $\mathbb{T}$ represents the set of groups of traders.

Condition (1) states that, taking other traders' payoff as given, a trader chooses his trading partner optimally. Hence, if a type z trader is paired with a type z' trader, he cannot obtain a higher level of utility from trading with someone else. It is effectively the standard stability condition.

2.3 Characterization of sorting on volatility

In this subsection, we focus on characterizing sorting based on the volatility type of traders. So, we assume that the volatility of traders is the only observable heterogeneity, $\mathbb{Z} = \sum \cup \{\emptyset\}$, and every trader has a units of asset. And we assume that correlation between the preference of a trader and that of his counterparty is always characterised $p \equiv \Pr(x^i x^j < 0) = \frac{1}{2}$. We will rationalise the correlation by introducing an additional heterogeneity in the next section. Our characterization follows two steps: First, we establish the necessary condition on the equilibrium allocation with Proposition 1. It shows that traders with relatively high volatility ($\sigma > \sigma^*$) must match with traders with relatively lower volatility ($\sigma \leq \sigma^*$). That is, the market must be partitioned in two subsets, divided by an endogenous cutoff type, σ^*. With this feature, we then characterize the equilibrium and show that the equilibrium payoff function is unique.

Lemma 1 is the key to understand the matching outcome in this economy. It shows that, the total surplus is always higher by pairing volatile types with stable types instead of letting volatile types match between themselves. This is true as long as preferences shocks are not *perfectly* negatively correlated. One can understand this by considering the simple case where there are only two types, $\sigma_L = 0$ and $\sigma_H > 0$. Whenever the preferences are not perfectly negatively correlated, there is always some probability that the extreme types end up being the same types. Whenever this happens, no surplus is generated if extreme types are paired together. On the other hand, by pairing an extreme type with a stable type, we effectively remove such risk. No matter what the realization is, the maximum aggregate surplus can be reached.

Lemma 2 *Let $\sigma_4 \geq \sigma_3 > \sigma_2 \geq \sigma_1$, for any $p < 1$,*

$$\Omega(\sigma_4, \sigma_3) + \Omega(\sigma_2, \sigma_1) < \Omega(\sigma_4, \sigma_1) + \Omega(\sigma_3, \sigma_2) = \Omega(\sigma_4, \sigma_2) + \Omega(\sigma_3, \sigma_1)$$

Proof. Clearly, for any contract that maximizes the surplus must satisfy the following condition: $a(\sigma, s) = 2a$ if $\varepsilon_\sigma > \varepsilon_{\sigma'}$ and $a(\sigma, s) = 0$ if $\varepsilon_\sigma < \varepsilon_{\sigma'}$. The trading surplus for each state s is then $z(s) \equiv |\varepsilon_\sigma - \varepsilon_{\sigma'}|a$. Hence, $\Omega(\sigma, \sigma') = E_s[z(s)] + \{W^0(\sigma) + W^0(\sigma')$ has the following expression:

$$\begin{aligned} \Omega(\sigma, \sigma') &= a\left[p(\sigma' + \sigma) + (1-p)|\sigma' - \sigma|\right] + W^0(\sigma) + W^0(\sigma') && (2) \\ &= a\left[\max(\sigma, \sigma') - (1-2p)\min(\sigma, \sigma')\right] + W^0(\sigma) + W^0(\sigma'). && (3) \end{aligned}$$

Therefore,

$$\begin{aligned} & [\Omega(\sigma_4, \sigma_3) + \Omega(\sigma_2, \sigma_1)] - [\Omega(\sigma_4, \sigma_1) + \Omega(\sigma_3, \sigma_2)] \\ = \; & a\left[(\sigma_4 + \sigma_2) - (1-2p)(\sigma_3 + \sigma_1)\right] - a\left[(\sigma_4 + \sigma_3) - (1-2p)(\sigma_1 + \sigma_2)\right] \\ = \; & a\left[-(\sigma_3 - \sigma_2) - (1-2p)(\sigma_3 - \sigma_2)\right] = -2(1-p)(\sigma_3 - \sigma_2) < 0 \end{aligned}$$

■

Intuitively, the stable type here acts as a "market maker" who is always ready to take the opposite position of "customers". Although the market maker himself might not need to trade, and despite that the customers can reach a higher pair-wise surplus with another customers, trading through market makers minimizes the uncertainty of the preference shocks in the economy. Hence, it is always efficient to match a marker maker with a customer.

As any competitive equilibrium must support the efficient matching, Proposition 1 then establishes the necessary condition for the matching outcomes.

Proposition 3 *Any equilibrium allocation f must satisfy the following conditions: if $f(\sigma, \sigma') > 0$ and $f(\hat{\sigma}, \hat{\sigma}') > 0$,*

$$\max(\sigma, \sigma') + \max(\hat{\sigma}, \hat{\sigma}') = \sigma_3 + \sigma_4 \qquad (4)$$

where σ_i is the ith order statistic of $\{\sigma, \sigma', \hat{\sigma}, \hat{\sigma}'\}$

Proof. Suppose not, consider an equilibrium where $f(\sigma_3, \sigma_4) > 0$ and $f(\sigma_2, \sigma_1) > 0$. Note that (1) can be rewritten as: $W^*(\sigma) + W^*(\sigma') \geq \Omega(\sigma, \sigma')$ for $\forall(\sigma, \sigma')$. Hence, we have $W^*(\sigma_4) + W^*(\sigma_2) \geq \Omega(\sigma_4, \sigma_2)$ and $W^*(\sigma_3)+W^*(\sigma_1) \geq \Omega(\sigma_3, \sigma_1)$, which implies $\Sigma W^*(\sigma_j) \geq \Omega(\sigma_4, \sigma_2)+\Omega(\sigma_3, \sigma_1)$. However, since $f(\sigma_3, \sigma_4) > 0$ and $f(\sigma_2, \sigma_1) > 0$ implies that $W^*(\sigma_4) + W^*(\sigma_3) = \Omega(\sigma_4, \sigma_3)$ and $W^*(\sigma_2) + W^*(\sigma_1) = \Omega(\sigma_1, \sigma_2) \implies \Sigma W^*(\sigma_j) = \Omega(\sigma_4, \sigma_3) + \Omega(\sigma_1, \sigma_2) > \Omega(\sigma_4, \sigma_2) + \Omega(\sigma_3, \sigma_1)$. Contradiction by Lemma 1. ■

Corollary 4 *There exists $\sigma^* \in [\sigma_L, \sigma_H]$ such that $f(\sigma, \sigma') = 0$ for each $(\sigma, \sigma') \in \Sigma_C \times \Sigma_C$ and $(\sigma, \sigma') \in \Sigma_M \times \Sigma_M$, where $\Sigma_M = [0, \sigma^*]$ and $\Sigma_C = [\sigma^*, \sigma_H]$.*

According to the above corollary, we can effectively solve the equilibrium as the standard assignment model with two-sided market. For convenience, we call traders with higher volatility $\sigma \in \Sigma_C$ "customers", as they are the ones who higher needs for trade. And we call traders with lower volatility $\sigma \in \Sigma_M$ "market-makers", as they are the ones who has lower need for trade ex-ante but always provide liquidity for "customers" in equilibrium. Those features are in fact the standard assumptions in the literature. Our key contribution is to show how such roles arise endogenously in equilibrium. As a result, we can further analyze the equilibrium payoff as "customers" and "market makers". In other words, how the surplus is split between "customers" and "market makers"? And what is the profit of market makers by providing such liquidity?

First of all, as we know from the two-sided market, complimentary determines the sorting. With corollary 1, (2) can be conveniently rewritten as

$$\Omega(\sigma_c, \sigma_m) = a\left[\sigma_c - (1-2p)\sigma_m\right] + W^0(\sigma_c) + W^0(\sigma_m) \tag{5}$$

where $\sigma_c \in \Sigma_C$ and $\sigma_m \in \Sigma_M$. Clearly, the additive nature of the payoff implies that there is no complimentary between "customers" and "market makers". In other words, as long as "customers" trade with "market makers", it should not matter which market maker they choose in equilibrium. The economy here is always clear: the value of market maker is generated by his ability to provide insurance. Such value is the same conditional on $\sigma_m < \sigma_c$.

Proposition 5 *For any $p > 1$, an unique equilibrium payoff $W^*(\sigma)$ is given by*

$$\begin{aligned} W^*(\sigma) &= \begin{cases} W(\sigma^*) + (1-2p)a(\sigma^* - \sigma), & \forall \sigma \in [0, \sigma^*] \\ W(\sigma^*) + (\sigma - \sigma^*)a, & \forall \sigma \in (\sigma^*, \sigma_H] \end{cases} \\ W(\sigma^*) &= \frac{a}{2}\{1 - (1-2p)\}\sigma^* + W^0(\sigma^*) = ap\sigma^* + ca \end{aligned}$$

where $W(\sigma^*) = ap\sigma^* + ca$ and σ^* solves

$$\int_0^{\sigma^*} dG(\tilde{\sigma}) = \int_{\sigma^*}^{\sigma_h} dG(\tilde{\sigma}).$$

Proof. See Appendix. ■

The above construction shares the standard property in the assignment model: the additional payoff gained by traderσ is exactly his contribution to the surplus within the match: $W_{\sigma}^{*}(\sigma) = \Omega_{\sigma}(\sigma, \tilde{\sigma})$ given he matches with $\tilde{\sigma}$ in equilibrium. In particular, from equation (5) ,conditional on a "customer" σ_c matches with a "market-maker" σ_m, the marginal contribution of a customer is given by $\Omega_{\sigma_c}(\sigma_c, \sigma_m) = a$; while the marginal contribution of a dealer is represented by $\Omega_{\sigma_m}(\sigma_c, \sigma_m) = -(1-2p)$. This then explains the shape of the equilibrium payoff function $W^{*}(\sigma)$.

2.4 Correlation of preference across traders

In this subsection, we rationalise the correlation of volatility of preference across agents by introducing an additional dimension of observable heterogeneity. We maintain the assumption that every agent is endowed with a units of asset. Assume that there are two groups $k \in \{A, B\}$ for any σ, each of size $g(\sigma)/2$. Let ε_{σ}^{i}denote the value of holding one unit of asset for trader with volatility σ in group k, which is given by

$$\varepsilon_{\sigma}^{i} = c + x_k^i \sigma,$$

and x_k^i is given by

$$\begin{aligned} x_k^i &= l_k \bar{X} \text{ with Prob } q \\ &= X_i \text{ with Prob } 1-q. \end{aligned}$$

where $\bar{X}$ and X_i are *uncorrelated* random variables that take value $\{-1, 1\}$ with equal probability, $l_A = 1$, and $l_B = -1$. One can think of $\bar{X}$ as the aggregate shock. l_k determines whether the correlation between group k's liquidity need and the aggregate shock is positive or negative. The probability q is the exposure to the aggregate shock. It also controls the correlation within group, which is $\rho_0(q) \equiv 1 - 2p_0$, where $p_0 \equiv \Pr(x_k^i x_k^j < 0) = (1-q)^2 \frac{1}{2}$ for $\forall\ j \neq i$. So, the correlation is an increasing function of q. The maximum and minimum correlation within the group is given by $\rho_0(1) = 1$ and $\rho_0(0) = 0$, respectively. On the other hand, by construction of l_A and l_B, the correlation across groups is then a decreasing function of q. More importantly, the cross-group correlation is always (weakly) smaller than within group correlation since $p_1 \equiv \Pr(x_k^i x_{k^{-1}}^j < 0) = q^2 + (1-q^2)\frac{1}{2} \geq p_0$ for $\forall q$. The equality holds when $q = 0$, which is the special case that all traders' preference are i.i.d. in our basic model. By construction, choosing groups is essentially a choice of the correlation. When $p_1 = 1$ (which happens when $q = 1$), the preference across groups are perfectly negatively correlated. This is then the environment where information is perfect: all traders know who the best counterparty is.

The constructed environment have two dimensional heterogeneity: demand volatility and correlation. Each trader is indexed by $z = (\sigma, k) \in \mathbb{Z} \equiv \Sigma \cup \{\emptyset\} \times \{A, B\}$. The trading surplus is similar as before expect that the correlation depends on whether the match is within-group or cross-group,

$$\Omega(z, \tilde{z}) = a \left[\max(\sigma, \tilde{\sigma}) - (1 - 2p_{k\tilde{k}}) \min(\sigma, \tilde{\sigma})\right] + W^0(\sigma) + W^0(\tilde{\sigma}),$$

where $p_{k\tilde{k}} = p_0$ when $k = \tilde{k}$ and $p_{k\tilde{k}} = p_1$ otherwise. The equilibrium definition is the same as before: a type z trader chooses his trading partner optimally based on the volatility and correlation.

To solve for this two-dimensional sorting problem, we first use the following Proposition to establish that traders must match with traders from the other group. The intuition is straightforward, as the trading surplus is higher when agents' preference are more negatively correlated. The problem is then reduced to the one dimensional sorting on volatility established in our basic model, where the effective correlation in equilibrium is given by the cross-group correlation.

Proposition 6 *Any equilibrium allocation f must satisfy Proposition 1 and $f((\sigma, k), (\tilde{\sigma}, k)) = 0$, $\forall \sigma, \tilde{\sigma} \in \sum$.*

Proof. See Appendix. ■

According to Proposition 6, the equilibrium allocation can be effectively reduced to our basic model by solving $f(\sigma_k, \sigma'_{-k})$ and setting the correlation parameter to be $p = \max\{p_0.p_1\} = p_1$. As the matching is always across groups, we can focus on the sorting on volatility. Hence, this two-dimensional setting can be easily characterized by the same method in the previous subsection, and their equilibrium payoff is then given by the following Proposition. For convenience, we can think of $\mathbb{Z} = \sum \cup \{\emptyset\}$.

Proposition 7 *For any $p > 1$, an unique equilibrium payoff $W^*(\sigma_k)$ is given by*

$$W^*(\sigma_k) = W^*(\sigma), \ \forall k \in \{A, B\},$$

where $W^*(\sigma)$ is defined in Proposition 5 by setting $p = p_1$.

2.5 Implementation of the general contract and bid-ask spread

A general contract, $\tau = \{(a(z, s), t(z, s)), a(z', s), t(z', s))\}$, specifies allocation of asset and transfer of numeraire goods. In general, the numeraire good transfer depends on the state s. Since we are talking about a market with spot transactions. In this subsection, we implement a general contract by a spot transaction contract, which specifies the price of transaction for each unit of asset transfer and total trade volume. As we can interpret one trader in a transaction as the market maker, with $\sigma \leq \sigma^*$, we call the price bid price, denoted by q^b, if the market maker buys from his customer, with $\sigma > \sigma^*$, and ask price, denoted by q^a, if the market make sells to his customer.[4] The equilibrium payoff can be simply implemented by market post a bid price and an ask price to all customers. And the bid-ask spread is proportional to the market maker's

[4]In general, the price could be a function of trade volume. In our current setup, it is only a function of direction of trade.

profit from trade. Denote the volatility type of a market marker to be σ_m and his asset holding a_m. Denote the volatility type of a customer to be σ_c and his asset holding a_c. So the spot transaction contract between a market maker of type $z_m = (\sigma_m, a_m)$ and a customer of type $z_c = (\sigma_c, a_c)$ is

$$\begin{aligned}
\{a(z_m, s), t(\sigma_m, s)\} &= \left\{a(z_m, s),\ -q^b \left[a(z_m, s) - a_m\right]\right\}, \text{ when } \varepsilon_{\sigma_c} < \varepsilon_{\sigma_m}, \\
\{a(z_m, s), t(\sigma_m, s)\} &= \left\{a(z_m, s),\ q^a \left[a_m - a(z_m, s)\right]\right\}, \text{ when } \varepsilon_{\sigma_c} > \varepsilon_{\sigma_m}.
\end{aligned}$$

In the basic model, every trader has a units of asset. So the trade volume between a market maker and a customer is always a.

$$\begin{aligned}
W(\sigma_m) &= \frac{1}{2} E(\tilde{\varepsilon} | \varepsilon_{\sigma_m} > \varepsilon_{\sigma_c}) \left(2a - q^b a\right) + \frac{1}{2} q^a a \\
&= \left[c + (p - (1-p))\, \sigma\right] a + a \frac{q^a - q^b}{2} \\
&= ca - (1 - 2p) \sigma a + a \frac{q^a - q^b}{2} \\
W(\sigma_c) &= \frac{1}{2} q^b a + \frac{1}{2} \left[(c + \sigma) 2a - q^a a\right] \\
&= (c + \sigma) a - a \frac{q^a - q^b}{2}
\end{aligned}$$

According to Proposition 5, $\frac{q^a - q^b}{2} = (1-p)\sigma^*$.

3 Dynamic model – endogenous link formation

In this section, we extend the basic model to a dynamic setting. This allows traders to form multiple links, make participation choices depending on their asset position. As a result, a richer network structure emerge in equilibrium, where traders' incentive to make market plays a crucial role.

3.1 Setup and equilibrium

The economy lasts $N + 1$ periods ($t = 0, 1, \ldots, N$). In period 0, a trader's trading partner for each period, from period 1 to period N are determined up to the volatility type, σ, asset holding, and which group the agent is in. So $\mathbb{Z} = \sum \cup \{\emptyset\} \times [0, 2a] \times \{A, B\}$.

Definition 8 *An equilibrium is an allocation function:* $f_t : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}_+$*, and equilibrium payoff* $W_t^*(\cdot) : \mathbb{Z} \to \mathbb{R}_+$ *for* $t = 1, \ldots, N$*, such that the following conditions are satisfied*

(1) Optimality for traders:

$$W_t^*(z) \geq \max_{\tilde{z} \in \mathbb{Z}} \Omega_t(z, \tilde{z}) - W_t^*(\tilde{z}), \tag{6}$$

and for any $f_t(z, z') > 0$*,* $z' \in \arg\max_{z \in \mathbb{Z}} \{\Omega_t(z, z') - W_t^*(z)\}$.

(2) Feasibility:

$$\int_{z\in\mathbb{Z}} f_t(z,\tilde{z})d\tilde{z} = h_t(z), \forall z \in \mathbb{Z}, t, \tag{7}$$

where $h_t(z)$ is the density function of z.

Equilibrium condition (6) is a sequential rationality condition. One of the key decisions in the matching problem is whether a trader will participate in trading in period t.

3.2 Characterisation

Group A traders will always choose to match with group B traders. And traders with assets will always choose to match with traders with no asset, all other observable characteristics given. Because of linearity in traders payoff, traders will always hold either 0 or $2a$ units of asset for any period $t > 1$. Given these implicit knowledge, we can simplify the state variable to be σ, and the state space, $\mathbb{Z}$, to $\sum \cup \{\emptyset\}$.

Proposition 9 *For an economy with N rounds of trade, the dynamic equilibrium follows a recursive structure, where matching at period t is characterised by a cutoff volatility type, σ_t^*, such that $G(\sigma_t^*) = \frac{1}{2^t}$, for $t = 1, \ldots, N$.*

- *Traders participating in period-t trade with $\sigma > \sigma_t^*$ match with traders participating in period-t trade with $\sigma \leq \sigma_t^*$.*
- *Traders with σ between σ_{t+1}^* and σ_t^* will participate in each round of trade until period $t+1$, for all $t = 1, \ldots, N$.*[5]
- *The equilibrium distribution is characterised by equations (8), (9) and (10). The equilibrium payoff of traders is determined by equations (11) ∼(16).*

Proof. See Appendix. ■

Equilibrium distribution For clarity of presentation, we assume that the preference of traders is not correlated, $p = \frac{1}{2}$. The proof for the proposition applies to arbitrary correlation.

The equilibrium distribution is summarised by cutoff types, σ_t^*, such that $G(\sigma_t^*) = 2^{-t}$, $\forall t \in \{1, \ldots, N\}$.

$$\int_{\sigma_t^*}^{\sigma_{t-1}^*} f_t(\sigma,\tilde{\sigma})d\tilde{\sigma} = g(\sigma), \quad \text{if} \quad \sigma_L \leq \sigma \leq \sigma_t^*, \tag{8}$$

$$\int_{\sigma_L}^{\sigma_t^*} f_t(\sigma,\tilde{\sigma})d\tilde{\sigma} = g(\sigma), \quad \text{if} \quad \sigma_t^* < \sigma \leq \sigma_{t-1}^*, \tag{9}$$

$$f_t(\sigma,\{\emptyset\}) = g(\sigma), \quad \text{if} \quad \sigma_{t-1}^* < \sigma \leq \sigma_H, \tag{10}$$

where $\sigma_0^* = \sigma_H$.

[5]Define $\sigma_0^* = \sigma_L$ and $\sigma_{N+1}^* = \sigma_H$.

Equilibrium payoff The equilibrium payoff from the trade in the last period can be determined in a way similar to that in the static game. The only difference is that depending on traders' preference type, asset holding, and whether they are dealers or not, they may choose not to participate in the trading game. This will affect the expected transfer from customers to dealers.

$$\begin{aligned} W_N^*(\sigma) &= \begin{cases} ca + T_N, & \forall \sigma \le \sigma_N^*, \\ \frac{1}{2}(c-\sigma)0 + \frac{1}{2}(c+\sigma)2a - T_N, & \forall \sigma_N^* < \sigma. \end{cases} \\ &= \begin{cases} ca + T_N, & \forall \sigma \le \sigma_N^*, \\ (c+\sigma)\, a - T_N, & \forall \sigma_N^* < \sigma. \end{cases} \end{aligned} \tag{11}$$

The transfer,

$$T_N = \frac{\sigma_N^*}{2} a, \tag{12}$$

from customers to dealers is pinned down by the indifference condition for an agent of σ_N^* type to be indifferent between being a market maker and being a customer. The payoff for a type σ_N^* dealer is $ca + T_N$. The payoff for a type σ_N^* to be a customer, is $(c + \sigma_N^*)a - T_N$.

Similarly, we can write down recursively traders' payoff at period $N-1$,

$$W_{N-1}^*(\sigma) = \begin{cases} W_N^*(\sigma) + T_{N-1}, & \forall \sigma \le \sigma_{N-1}^*, \\ (c+\sigma)\, a - T_{N-1}, & \forall \sigma_{N-1}^* < \sigma. \end{cases} \tag{13}$$

To pin down, the transfer to dealers, T_{N-1}, we have to check the participation decision of traders. Clearly market makers always have incentive to participate in the trading game. Trading does not affect their expected asset holding but gives them additional transfers. The payoff for type σ_{N-1}^* to be a dealer is, $W_N^*(\sigma_{N-1}^*) + T_{N-1}$. As $\sigma_{N-1}^* \ge \sigma_N^*$, $W_N^*(\sigma_{N-1}^*) = (c+\sigma)\, a - T_N$. The payoff for a type σ_{N-1}^* to be a customer, is $(c+\sigma)\, a - T_{N-1}$, because period $N-1$ customers will leave the market after the trade. Therefore,

$$T_{N-1} = \frac{T_N}{2}. \tag{14}$$

More generally, for any $1 < t < N$,

$$W_t^*(\sigma) = \begin{cases} W_{t+1}^*(\sigma) + T_t, & \forall \sigma \le \sigma_t^*, \\ (c+\sigma)a - T_t, & \forall \sigma_t^* < \sigma. \end{cases} \tag{15}$$

As, $\sigma_t^* \ge \sigma_{t+1}^*$, the payoff for type σ_t^* to be a dealer is, $W_t^*(\sigma_t^*) + T_t = (c+\sigma_t^*)\, a - T_{t+1} + T_t$. The payoff for a type σ_t^* to be a customer, is $(c+\sigma_t^*)\, a - T_t$. $T_t = \frac{T_{t+1}}{2}$.

For an N-period trading game, the profit from market making for those agents with $\sigma < \sigma_N^*$, is

$\sum_{n=1}^{N} \frac{1}{2^{N-n}} T_N = 2T_N(1 - \frac{1}{2^N}) = \sigma_N^*(1 - 2^{-N})a$. The payoff at period 0 is,

$$W_0^*(\sigma) = \begin{cases} \left(c + \sigma - 2^{-N}\sigma_N^*\right) a, & \forall \sigma \in [\sigma_1^*, \bar{\sigma}] \\ (c+\sigma)a + \sum_{\tau=1}^{t-1} T_\tau - T_t, & \forall \sigma \in [\sigma_t^*, \sigma_{t-1}^*], t \in \{2, \ldots, N\} \\ \left[c + \left(1 - 2^{-N}\right)\sigma_N^*\right] a, & \forall \sigma \in [0, \sigma_N^*] \end{cases}$$

$$= \begin{cases} \left(c + \sigma - 2^{-N}\sigma_N^*\right) a & \forall \sigma \in [\sigma_N^*, \bar{\sigma}] \\ \left[c + \left(1 - 2^{-N}\right)\sigma_N^*\right] a, & \forall \sigma \in [0, \sigma_N^*] \end{cases} \tag{16}$$

Corollary 10 *As $N \to \infty$, the equilibrium asset allocation and equilibrium payoff converges to those in a competitive equilibrium with market clear price c.*

4 Implications of the dynamic model

In this section, we study the implications of the dynamic model on the property of the endogenous network structure, and market liquidity, measured by trade volume, bid-ask spread, and the total number of trading rounds. To keep presentation simple, we assume there is no correlation between the volatility types of traders. As we know from the analysis in previous sections, the model implication is qualitatively the same with non-trivial correlation.

4.1 Distribution of links and measure of the network structure

The participation decision will determine the equilibrium number of trading links a trader has after N rounds of trading. Denote the participation decision $\pi_t : \mathbb{Z} \to [0, 1]$. $\pi_t(z) = 0$ if the trader does not participate in period-t trade.

$$\pi_t(\sigma) = \begin{cases} 0, & \forall \sigma \in [\sigma_{t-1}^*, \sigma_H], \\ 1, & \forall \sigma \in [\sigma_L, \sigma_{t-1}^*]. \end{cases}$$

Define $l(\sigma) = \sup\{t : \ \pi_t(\sigma) > 0, \ t \in \{1, \ldots, N\}\}$. With N-period trade, the number of links a trader has is

$$l(\sigma) = \begin{cases} 1, & \forall \sigma \in [\sigma_1^*, \bar{\sigma}], \\ t - 1, & \forall \sigma \in [\sigma_t^*, \sigma_{t-1}^*], t \in \{2, \ldots, N\} \ , \\ N, & \forall \sigma \in [\sigma_L, \sigma_N^*]. \end{cases}$$

which implies that more efficient market makers, those with smaller σ, are more active in the market. They not only trade with customers in period 1, but also offer intermediation services to less efficient market makers, who have less links in equilibrium. This is consistent with findings in Neklyudov (2013)[18]. The more efficient market makers are also more central.

Another observation is that as the market becomes more liquid (higher N), the distribution of trading links becomes more concentrated with more efficient market makers. In figure 1, we illustrate the network structure by comparing the network when $N = 1$ and when $N = 7$. We locate traders of lower volatility type closer to the centre of the graph.

So, the distribution of the number of links in the market follows an exponential distribution.

$$\Pr(l = n) = \begin{cases} \frac{1}{2^l} & \text{if } l = 1, \ldots, N-1, \\ \frac{1}{2^{N-1}} & \text{if } l = N. \end{cases}$$

In empirical works, the distribution of the number of links in an OTC market typically satisfies power law (see for example [7]). The distribution from our still stylised model does not have the thick tail property. It remains an open question to us what extension is needed to capture the thick tailed distribution in reality. In figure 2, we illustrate the distribution of trading links with the case $N = 10$.

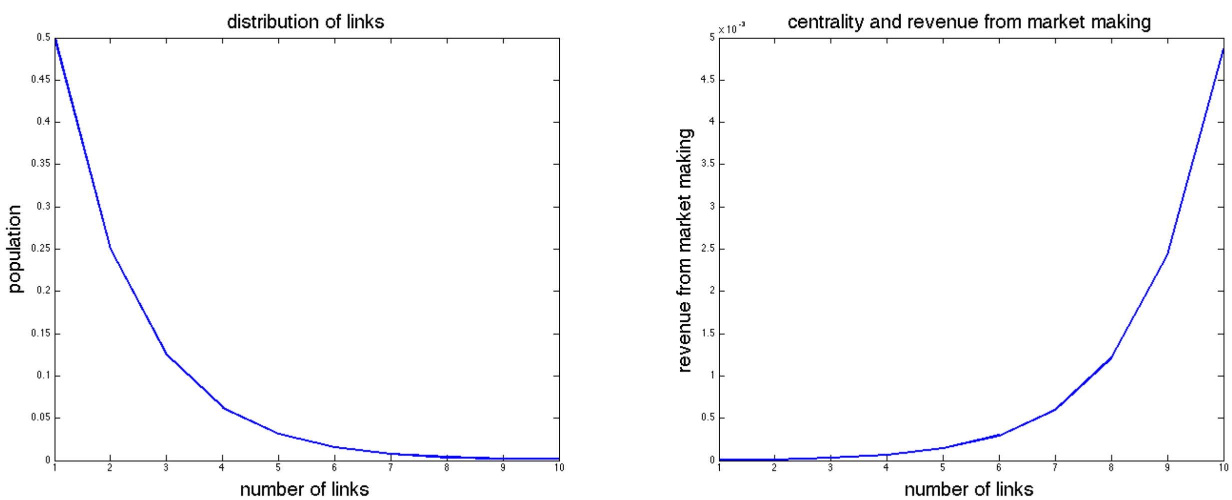


Figure 2: The distribution of trading links, revenue from market making ($N = 10$).

Given the distribution of links, we could measure the completeness of the network by looking at the average number of links. The average number of links with N-period trade, $\bar{l}$, is

$$\bar{l}(N) = \sum_{i=1}^{N} \frac{i}{2^i} + \frac{N}{2^N}.$$

$\bar{l}$ is an increasing function in N.A network, with N-round trade, has at most N links on average. So we characterise the *sparsity* of network by $\psi(N) \equiv \frac{\bar{l}(N)}{N}$.

$$\psi(N) = \sum_{i=1}^{N} \frac{i/N}{2^i} + \frac{1}{2^N}.$$

As $\psi(N)$ is strictly decreasing in N, the network becomes more sparse at N increases. At the limit $N \to \infty$, where the allocation is the same as in a competitive market, $\psi(\infty) = 0$.

4.2 Trade volume

The total expected trade volume, $\mathbb{E}Vol(\sigma)$, conditional on type σ, is

$$\mathbb{E}Vol(\sigma) = \begin{cases} a, & \forall \sigma \in [\sigma_1^*, \sigma_H], \\ a(t-1), & \forall \sigma \in [\sigma_t^*, \sigma_{t-1}^*], t \in \{2, \ldots, N\} \\ aN, & \forall \sigma \in [\sigma_L, \sigma_N^*]. \end{cases},$$

So, the total trade volume of the economy is increasing in N. But, the marginal increase in trade volume is decreasing in N.

Figure 3 illustrates how the gross trade volume and net trade volume depend on the preference type of the trader. Traders with preference type close to c has higher gross trade volume but less net trade volume, as a percentage of the gross trade volume. This shows clearly the market-making function of traders with less volatile preference.

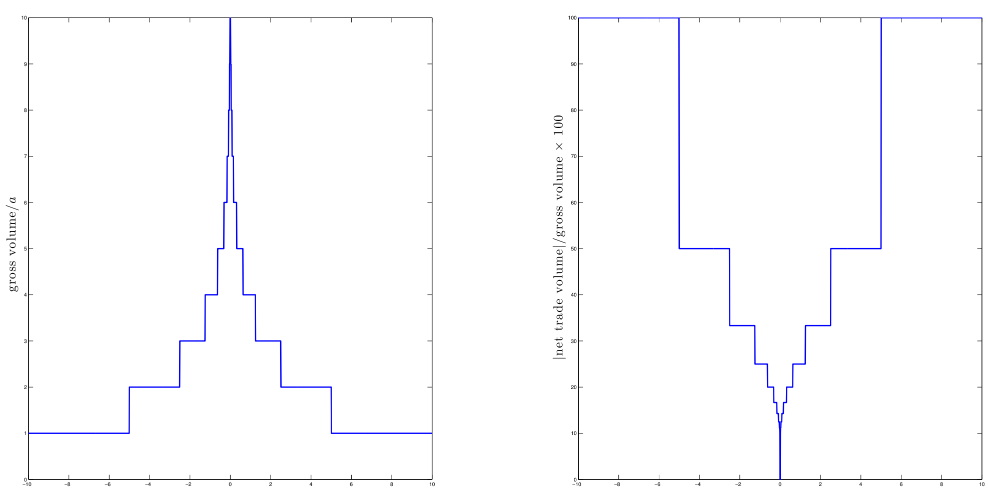


Figure 3: Trade volume across the preference type of traders. ($N = 10$)

4.3 Revenue from market-making and bid-ask spread

The revenue from market-making, $R(\sigma)$, depends on the volatility type of the market maker,

$$R(\sigma) = \begin{cases} 0, & \forall \sigma \in [\sigma_1^*, \sigma_H] \\ \left(2^{-(N-t+1)} - 2^{-N}\right) \sigma_N^* a, & \forall \sigma \in [\sigma_t^*, \sigma_{t-1}^*], t \in \{2, \ldots, N\} \\ \left(1 - 2^{-N}\right) \sigma_N^* a, & \forall \sigma \in [\sigma_L, \sigma_N^*] \end{cases} .$$

A more efficient market maker makes more profit from market making, as he is more central in the network. See figure 2 for an illustration. The payment as customer, $T(\sigma)$, also depends on the type of the customer, as different types act as customers at different moment,

$$T(\sigma) = \begin{cases} 2^{-N} \sigma_N^* a, & \forall \sigma \in [\sigma_1^*, \sigma_H] \\ 2^{-N+t-1} \sigma_N^* a, & \forall \sigma \in [\sigma_t^*, \sigma_{t-1}^*], t \in \{2, \ldots, N\} \\ 0, & \forall \sigma \in [\sigma_L, \sigma_N^*] \end{cases} .$$

So the expected revenue of market makers, per unit of asset transaction, is

$$\pi(t) = 2^{-N+t-2} \sigma_N^*, \forall t \in \{1, \ldots, N\},$$

which implies that the bid-ask spread at period t is

$$q_t^a - q_t^b = 2^{-N+t-1} \sigma_N^*, \forall t \in \{1, \ldots, N\}.$$

Notice that the bid-ask spread is increasing in t, which means it is more costly to trade with more efficient dealers. This is consistent with findings in Li and Schürhoff (2014) [17].

4.4 Endogenous market liquidity

The marginal effect of total trading rounds, N, on total trade volume and allocation, is diminishing. If it is costly for market makers to create liquidity by staying for more rounds of transactions, the market liquidity, measured by N, can be endogenous. With entry cost F into the market, the total number of trading rounds in equilibrium is such that all traders are weakly better off participating in the game. So, the equilibrium trading round, $N^*(F)$, solves the following problem,

$$N^*(F) = \max\{n : \sigma_n^* (1 - 2^{-n}) a > F\}$$

$N^*(F)$ is a decreasing function in the entry cost F. When the volatility distribution is more dispersed, so that σ_n^* is higher for any n, $N^*(F)$ is larger. This implies that the market is more liquid, when the economy is more volatile so there is more need to trade. This is consistent with findings in Peltonen, Scheicher and Vuillemey(2014)[19].

5 Conclusion

To conclude, we build a dynamic model of over-the-counter market, where market making activities and a network structure emerge endogenously. The network structure is qualitatively similar to the network structure we observe in a typical OTC market. The key mechanism behind these results is the sorting on the volatility of traders' preference over the asset. Market making services offered by traders with less volatile preference insure traders with more volatile preference against their trading needs, which could be either selling or buying the asset. The model gives us a fresh understanding of the economics behind the trading patterns in the OTC market. The policy implication of the model on the risk and stability of the trading structure in the OTC market may therefore be very different from that in the existing literature.

References

[1] Gara Afonso and Ricardo Lagos. An empirical study of trade dynamics in the fed funds market. *FRB of New York staff report*, (550), 2012.

[2] Gara Afonso and Ricardo Lagos. Trade dynamics in the market for federal funds. Technical report, National Bureau of Economic Research, 2014.

[3] Gara Afonso, Anna Kovner, and Antoinette Schoar. Trading partners in the interbank lending market. *FRB of New York Staff Report*, (620), 2013.

[4] Andrew G Atkeson, Andrea L Eisfeldt, and Pierre-Olivier Weill. Entry and exit in otc derivatives markets. Technical report, National Bureau of Economic Research, 2014.

[5] Ana Babus. The formation of financial networks. 2007.

[6] Ana Babus. Endogenous intermediation in over-the-counter markets. *Available at SSRN 1985369*, 2012.

[7] Morten L Bech and Enghin Atalay. The topology of the federal funds market. *Physica A: Statistical Mechanics and its Applications*, 389(22):5223–5246, 2010.

[8] Jonathan Chiu and Cyril Monnet. Relationship lending in a tiered interbank market. working paper, 2014.

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

[10] Maryam Farboodi. Intermediation and voluntary exposure to counterparty risk. *Available at SSRN 2535900*, 2014.

[11] Douglas M Gale and Shachar Kariv. Financial networks. *The American Economic Review*, pages 99–103, 2007.

[12] Michael Gofman. Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *Available at SSRN 2194357*, 2014.

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

[14] Matthew O Jackson. A survey of network formation models: stability and efficiency. *Group Formation in Economics: Networks, Clubs, and Coalitions*, pages 11–49, 2005.

[15] Ricardo Lagos and Guillaume Rocheteau. Liquidity in asset markets with search frictions. *Econometrica*, 77(2):403–426, 2009.

[16] Benjamin Lester, Guillaume Rocheteau, and Pierre-Olivier Weill. Competing for order flow in otc markets. Technical report, National Bureau of Economic Research, 2014.

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

[18] Artem V. Neklyudov. Bid-ask spreads and the decentralized interdealer markets: Core and peripheral dealers. Technical report, Working Paper, University of Lausanne, 2013.

[19] Tuomas A Peltonen, Martin Scheicher, and Guillaume Vuillemey. The network structure of the cds market and its determinants. *Journal of Financial Stability*, 13:118–133, 2014.

[20] Ariel Rubinstein and Asher Wolinsky. Middlemen. *The Quarterly Journal of Economics*, pages 581–594, 1987.

A Appendix

A.1 Omitted Proof

A.1.1 Proof for Proposition 5

Define $W(\sigma,\sigma') \equiv \Omega(\sigma,\sigma') - W^*(\sigma')$. For any $\sigma' \in [\sigma^*, \sigma_H]$,

$$W(\sigma,\sigma') = \begin{cases} \sigma' a - (1-2p)\sigma - W^*(\sigma'), & \text{for } \sigma' > \sigma, \\ \sigma a - (1-2p)\sigma' - W^*(\sigma'), & \text{for } \sigma \geq \sigma' \geq \sigma^*, \\ \sigma a - (1-2p)\sigma' - W^*(\sigma'), & \text{for } \sigma^* > \sigma'. \end{cases}$$

Proof. By construction of $W^*(\sigma)$:

$$W_2(\sigma,\sigma') = \begin{cases} 0, & \text{for } \sigma' > \sigma, \\ -(1-2p)-1 = -2(1-p) < 0, & \text{for } \sigma \geq \sigma' \geq \sigma^*, \\ -(1-2p)-(1-2p) = 0, & \text{for } \sigma^* > \sigma'. \end{cases}$$

Moreover, for $\sigma' > \sigma > \sigma^* > \sigma''$: $W(\sigma,\sigma'') - W(\sigma,\sigma') = \sigma a - (1-2p)\sigma'' - W^*(\sigma'') - \{\sigma' a - (1-2p)\sigma - W^*(\sigma')\} = 2a(1-p)(\sigma-\sigma^*) > 0$. Hence, $[\sigma_L, \sigma^*] = \arg\max_{\sigma'} W(\sigma,\sigma')$ for any $\sigma' \in [\sigma^*, \sigma_H]$. Similarly, for any $\sigma \in [0, \sigma^*]$,

$$W_2(\sigma,\sigma') = \begin{cases} 0, & \text{for } \sigma' > \sigma, \\ 2(1-p) > 0, & \text{for } \sigma \geq \sigma' \geq \sigma^*, \\ 0, & \text{for } \sigma^* > \sigma'. \end{cases}$$

And for $\sigma' > \sigma^* > \sigma > \sigma''$: $W(\sigma,\sigma') - W(\sigma,\sigma'') = 2a(1-p)(\sigma^*-\sigma) > 0 \implies [\sigma^*, \sigma_H] = \arg\max_{\sigma'} W(\sigma,\sigma')$ for any $\sigma' \in [0,\sigma^*]$. Lastly, one can see that this payoff satisfies the feasible within each pair (i.e.,$W^*(\sigma) + W^*(\sigma') \leq \Omega(\sigma,\sigma')$):

$$\begin{aligned} W^*(\sigma_c) + W^*(\sigma_m) &= 2W(\sigma^*) + (1-2p)a(\sigma^*-\sigma_m) + (\sigma_c - \sigma^*)a \\ &= \Omega(\sigma_c,\sigma_m) \end{aligned}$$

■

A.1.2 Proof for Proposition 6

Proof. The logic is the same as before, we show that when either of the above conditions is violated, there is a surplus left and the aggregate surplus can therefore be improved by rearranging the match. For notational convenience, we use σ_k to denote type-(σ, k). First, consider the case that $f(\sigma_A, \sigma_A') > 0$ and $f(\tilde{\sigma}_B, \tilde{\sigma}_B') > 0$ and Proposition 3 is not satisfied: $\sigma_A \geq \sigma_A' > \tilde{\sigma}_B \geq \tilde{\sigma}_B'$

$$\begin{aligned} & \Omega(\sigma_A, \sigma_A') + \Omega(\tilde{\sigma}_B, \tilde{\sigma}_B') \\ = \;& a\left[\max\left(\sigma_A, \sigma_A'\right) - (1-2p_0)\min\left(\sigma_A, \sigma_A'\right)\right] \\ & + a\left[\max\left(\tilde{\sigma}_B, \tilde{\sigma}_B'\right) - (1-2p_0)\min\left(\tilde{\sigma}_B, \tilde{\sigma}_B'\right)\right] \\ \leq \;& a\left(\sigma_3 + \sigma_4\right) - (1-2p_0)\left(\sigma_1 + \sigma_2\right) \\ < \;& a\left(\sigma_3 + \sigma_4\right) - (1-2p_1)(\sigma_1 + \sigma_2) = \Omega\left(\sigma_A, \tilde{\sigma}_B'\right) + \Omega\left(\tilde{\sigma}_B, \sigma_A'\right), \end{aligned}$$

where σ_i is the ith order statistic of $\{\sigma_A, \sigma'_A, \tilde{\sigma}_B, \tilde{\sigma}'_B\}$. Second, suppose that Proposition 1 is satisfied –that is, $f(\sigma_k, \sigma'_{k'}) > 0$ only if $\sigma_k \subset [0, \sigma^*]$ and $\sigma'_{k'} \subset [\sigma^*, \sigma_H]$ – but $f(\sigma_A, \sigma'_A) > 0$ and $f(\tilde{\sigma}_B, \tilde{\sigma}'_B) > 0$,

$$\begin{aligned}
& \Omega(\sigma_A, \sigma'_A) + \Omega(\tilde{\sigma}_B, \tilde{\sigma}'_B) \\
= \; & a\left(\sigma'_A + \tilde{\sigma}'_B\right) - (1 - 2p_0)\left(\sigma_A + \tilde{\sigma}_B\right) \\
< \; & a\left(\sigma'_A + \tilde{\sigma}'_B\right) - (1 - 2p_1)\left(\sigma_A + \tilde{\sigma}_B\right) \\
< \; & \Omega\left(\sigma_A, \tilde{\sigma}_B\right) + \Omega\left(\sigma_A, \tilde{\sigma}'_B\right).
\end{aligned}$$

Lastly, consider the case that $f(\sigma_k, \sigma'_k) = 0$ (that is, traders only match within each group) but the proposition 1 is not satisfied. Lemma 1 can be applied directly to this case within each group k. Hence, an allocation f maximizes the aggregate surplus if and only if Proposition 3 and $f(\sigma_k, \sigma'_k) = 0$ are satisfied. ■

A.1.3 Proof for Proposition 7

Proof. As proved in Proposition 5, the constructed payoff guaranteed that, it's optimal for a more volatile type to match with a relatively stable type across groups. Hence, all we need to show now is that there is no profitable deviation for traders to match traders within the group:

$$\begin{aligned}
& \Omega(\sigma_k, \tilde{\sigma}_k) - W^*(\tilde{\sigma}_k) \\
= \; & a\left[\max(\sigma_k, \tilde{\sigma}_k) - (1 - 2p^-)\min(\sigma_k, \tilde{\sigma}_k)\right] - W^*(\tilde{\sigma}_k) \\
< \; & a\left[\max(\sigma_k, \tilde{\sigma}_{-k}) - (1 - 2p^+)\min(\sigma_k, \tilde{\sigma}_{-k})\right] - W^*(\tilde{\sigma}_{-k}) \\
= \; & \Omega(\sigma_k, \tilde{\sigma}_{-k}) - W^*(\tilde{\sigma}_{-k}) \leq W^*(\sigma_k).
\end{aligned}$$

■

A.1.4 Proof for Proposition 9

Proof. When there is a correlation, for general p :

$$\begin{aligned}
W_N(\sigma) &= ca - (1 - 2p)\sigma a + \frac{q^a - q^b}{2} \\
&= -(1 - 2p)\sigma a + ca + T_N,
\end{aligned}$$

$$W_N(\sigma^*) = (c + \sigma^*)a - T_N \Rightarrow T_N = (1 - p)\sigma_N^* a,$$
$$\frac{T_t}{\sigma_t} = \frac{T_N}{\sigma_N \varphi_t(2^{N-t})} = \frac{(1 - p)a}{\varphi_t(2^{N-t})},$$

where the relationship between σ_t and σ_N is given by φ_t, where $\varphi_t \leq 1$ and equality holds only when $t = N$.

To show that the assignment satisfies traders' optimality for t :

Traders in each period essentially have matching decisions and participation next period. First, consider traders $\sigma > \sigma_t^*$: Clearly, if trader-σ matches with "dealers" this period, he must match none next period, as there is no surplus left so no reason for him to pay for the spread next period. So, what is left to show is that there is no profitable deviation if (1) trader-σ matches any trader $\sigma' > \sigma_t^*$ and leave the market; nor (2) he matches any trader $\sigma' > \sigma_t^*$ and stay the market for another period. Consider the first case that both customers leave the market afterward, the trading surplus with any $\sigma' > \sigma_t^*$ is given by:

$$\Omega_t((\sigma, 0), (\sigma', 2a)) = (\max\{\sigma, \sigma'\} - (1-2p)\max\{\sigma, \sigma'\})\, a$$

where $W_t^*(\sigma') = (c+\sigma')a - T_t$ for $\sigma' \geq \sigma_t^*$

$$\begin{aligned}
W_t(\sigma|\sigma') &= \Omega_t(\sigma, \sigma') - W_t(\sigma') \\
&= \begin{cases} -(1-2p)\sigma + T_t + ca, & \text{if } \sigma' > \sigma > \sigma^*, \\ [\sigma - 2(1-p)\sigma']\, a + T_t + ca, & \text{if } \sigma > \sigma' > \sigma^*, \\ \sigma a - (1-2p)\sigma' a - [-(1-2p)\sigma' a + T_t] + ca, & \text{if } \sigma^* > \sigma', \end{cases} \\
&= \begin{cases} -(1-2p)\sigma a + T_t + ca, & \text{if } \sigma' > \sigma > \sigma^*, \\ \sigma a - 2(1-p)\sigma' a + T_t + ca, & \text{if } \sigma > \sigma' > \sigma^*, \\ (c+\sigma)a - T_t, & \text{if } \sigma^* > \sigma'. \end{cases}
\end{aligned}$$

Note that $\frac{d}{d\sigma'}\{W_t(\sigma|\sigma')\} = 0$ for $\sigma^* > \sigma'$ and $\frac{d}{d\sigma'}\{W_t(\sigma|\sigma')\} < 0$ for $\sigma' \in [\sigma, \sigma^*]$ and $\frac{d}{d\sigma'}\{W_t(\sigma|\sigma')\} = 0$ for $\sigma' > \sigma$. Furthermore, when $\sigma' > \sigma > \sigma_t^*$:

$$\begin{aligned}
W_t^*(\sigma) - W_t(\sigma|\sigma') &= W_t^*(\sigma) - [\Omega_t(\sigma, \sigma') - W^*(\sigma')] \\
&= W_t^*(\sigma) - [-(1-2p)\sigma a + T_t + ca] \\
&= 2(1-p)\sigma a - 2T_t \\
&= 2(1-p)(\sigma - \sigma_t^*)a + 2(1-p)\sigma_t^* - 2T_t \\
&= 2(1-p)(\sigma - \sigma_t^*)a + 2(1-p)\sigma_t^* - \frac{2(1-p)}{(2^{N-t})\varphi_t}\sigma_t^* a \\
&= 2(1-p)(\sigma - \sigma_t^*)a + 2(1-p)\sigma_t^* a \left[1 - \frac{1}{(2^{N-t})\varphi_t}\right] > 0
\end{aligned}$$

where

$$\frac{T_t}{\sigma_t} = \frac{T_N}{\sigma_N \varphi_t (2^{N-t})} = \frac{(1-p)a}{\varphi_t(2^{N-t})}$$

and the relationship between σ_t and σ_N is given by φ_t (which depends on the distribution), where $\varphi_t \geq 1$ and equality holds only when $t = N$. Hence, $[0, \sigma_t^*] = \arg\max_{\sigma'} W_t(\sigma, \sigma')$ for any $\sigma \geq \sigma_t^*$. Now, consider the second case, where trader-σ matches $\sigma' > \sigma^*$ and stays for another period and match with any trader $\sigma < \sigma_{t+1}^*$. In this case, traders' utility is given by $\Omega_t(\sigma, \sigma') - W^*(\sigma') + T_t - T_{t+1} < \Omega_t(\sigma, \sigma') - W^*(\sigma') < W^*(\sigma)$. Hence, it's optimal for trader$-\sigma$ to match $\sigma \leq \sigma_t^*$ and leave the market after t.

Second, concider the case for traders $\sigma < \sigma_t^*$. Clearly, if they match with customer this period, they're strictly better off if they stay. So, what is left to show is that there is no profitable deviation if they choose to match another trader $\sigma' < \sigma_t^*$.

(Case 1) Consider the case that one of the "dealer" is the customer next period. Suppose $\sigma_t^* > \sigma' > \sigma_{t+1}^* > \sigma$:

$$
\begin{aligned}
W_t\left((\sigma,0)|(\sigma',2)\right) &= \Omega\left((\sigma,0),(\sigma',2)\right) - W_t^*(\sigma',2) \\
&= \frac{1}{2}\left[W_{t+1}(\varepsilon_L,\sigma',0) + W_{t+1}(\tilde{\varepsilon},\sigma,2)\right] + \frac{1}{2}\left[W_{t+1}(\varepsilon_H,\sigma',2) + W_{t+1}(\tilde{\varepsilon},\sigma,0)\right] - W_t^*(\sigma',2) \\
&= 0 + W_{t+1}^*(\sigma) + \frac{1}{2}(c+\sigma')2a - \left[T_t + W_{t+1}^*(\sigma')\right] \\
&= (c+\sigma')a - T_t + W_{t+1}^*(\sigma) - W_{t+1}^*(\sigma') \\
&= (c+\sigma')a - T_t + W_{t+1}^*(\sigma) - \left[(c+\sigma')a - T_{t+1}\right] \\
&= -T_t + W_{t+1}^*(\sigma) + T_{t+1} \\
&= -T_t + W_{t+1}^*(\sigma) + 2T_t = W_t^*(\sigma)
\end{aligned}
$$

Same for $W_t((\sigma,2)|(\sigma',0)) = W_t^*(\sigma)$. Similariy, when $\sigma_t^* > \sigma > \sigma_{t+1}^* > \sigma'$:

$$
\begin{aligned}
W_t\left((\sigma,0)|(\sigma',2)\right) &= \Omega((\sigma,0),(\sigma',2)) - W_t^*(\sigma',2) \\
&= \frac{1}{2}\left[W_{t+1}(\varepsilon_L,\sigma,0) + W_{t+1}(\tilde{\varepsilon},\sigma',2)\right] + \frac{1}{2}\left[W_{t+1}(\varepsilon_H,\sigma,2) + W_{t+1}(\tilde{\varepsilon},\sigma',0)\right] - W_t^*(\sigma',2) \\
&= 0 + \frac{1}{2}(c+\sigma)2a + W_{t+1}^*(\sigma') - \left[T_t + W_{t+1}^*(\sigma')\right] \\
&= (c+\sigma)a - T_t = W_t(\sigma') = T_t + \left[(c+\sigma)a - T_{t+1}\right]
\end{aligned}
$$

(Case 2): Consider the case that both dealers are dealers next period: Suppose $\sigma_t^* > \sigma_{t+1}^* > \sigma' > \sigma$:

$$
W_t(\varepsilon,\sigma',a) = T_t + W_{t+1}(\varepsilon,\sigma')
$$

$$
\begin{aligned}
W_t((\sigma,0)|(\sigma',2)) &= \frac{1}{2}\left[W_{t+1}(\varepsilon_L,\sigma',0) + W_{t+1}(\tilde{\varepsilon},\sigma,2)\right] + \frac{1}{2}\left[W_{t+1}(\varepsilon_H,\sigma',2) + W_{t+1}(\tilde{\varepsilon},\sigma,0)\right] - W_t^*(\sigma',2) \\
&= W_{t+1}^*(\sigma') + W_{t+1}(\sigma) - \left[T_t + W_{t+1}^*(\sigma')\right] \\
&= -T_t + W_{t+1}(\sigma) < W_t^*(\sigma) = T_t + W_{t+1}(\sigma)
\end{aligned}
$$

Asset changing hands among dealers will not change their continuation value. Hence, they will loss by giving up $2T_t$. Hence, we have shown that constructed matching role as well as market particiaption decision satisfy traders' optimality condition. ■