

# Endogenous Market Making and Network Formation

Briana Chang\*

Shengxing Zhang<sup>†</sup>

University of Wisconsin–Madison

London School of Economics

April 28, 2015

## Abstract

This paper develops a dynamic model of network formation in over-the-counter markets. Traders choose with whom to connect as well as whether to remain active in each period. Even though all traders have the same trading technology, we show that traders with higher trading needs optimally choose to match with traders with lower needs for trade, and also leave the market earlier. The model thus endogenously generates a core-periphery market structure, where traders who do not need to trade turn out to be the most connected and have the highest gross trade volume. Since 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 why the financial architecture involves a small number of large, interconnected institutions. We use this framework to analyze how underlying frictions change bid-ask spreads, trading volume, and asset allocation. We further apply it to the market of interbank lending and study how the structure determines the extent of contagion.

*preliminary*

**Keyword:** Over-the-Counter Market, Core-Periphery Trading Network, Matching, Intermediation

**JEL classification:** C70, G12, G21

---

\*Email: bchang@bus.wisc.edu.

<sup>†</sup>Email: S.Zhang31@lse.ac.uk. Zhang thanks the Centre for Macroeconomics at LSE for their support.

# 1 Introduction

Trading in decentralized markets or over-the-counter (OTC) markets requires traders to find a counterparty. Empirically, we observe that these markets involve active intermediaries and exhibit a core-periphery structure, where certain traders intermediate a large amount of the trades (see, for example, Li and Schurhoff (2011)[24] and Bech and Atalay (2010)).<sup>1</sup> The fact that the financial architecture involves a small number of large, interconnected banks clearly has an important implication on the stability of financial system.<sup>2</sup> Yet, what remains unknown is why such trading structure arises in the first place. In other words, why do traders who have higher needs for trade (customers) choose to trade with traders who have fewer needs for trade (dealers)? And why do certain financial institutions become more connected and systemically important than others?

In order to understand how such structure is formed, we develop a dynamic matching model that allows traders to optimally choose their trading links for each period. This is the key difference from the existing search models on OTC markets (starting from Duffie, et al. (2005) [12]), where traders randomly match with their counterparty. Moreover, traders are also allowed to choose how many traders with whom to connect. The model thus generates an endogenous trading structure, where all trading links as well as the number of links for each trader are determined in equilibrium. It further gives predictions on which traders will become the most connected. Our model can therefore deliver features similar to those in network-based frameworks without imposing an exogenous structure. Since all trading links in our model are neither random nor exogenous, we can therefore explicitly answer why we observe such a trading structure in the market and analyze whether this outcome is efficient.

The key heterogeneity in our model is traders' exposure to preference shocks. Traders who are subject to more volatile shocks have more an 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. For example, in the interbank lending market, volatility represents the liquidity shortage that arises from deposit withdrawal or large payments from one bank to another. Some banks may have more volatile liquidity needs than others. This could be because they are small banks or have a less stable customer base. On the other hand, traders with more stable positions are the ones who gain less from participating in the market. We can think of these banks as the ones with more diversified portfolios (for example, larger banks) and therefore the one who are less vulnerable to such shocks. Hence, they do not need to borrow from or lend to other

---

<sup>1</sup>Li and Schurhoff (2011)[24] and Bech and Atalay (2010)[10] document the hierarchical core-peripheral structure in the municipal bond market and the federal funds market, respectively. They show that the number of dealer connections is heavily skewed with a fat right tail populated by several core dealers.

<sup>2</sup>A growing literature focuses on the role of financial networks as an amplification mechanism: for example, Allen et al. (2000)[5], Acemoglu et al. (2014)[1], and Gofman (2014) [16] study the financial contagion in interbank lending markets, where the network structure is given exogenously.

banks.

The main decision of traders in our model involves the formation of trading links. Before the realization of preference shocks, traders choose with whom to match each period, and they agree on the allocation and 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 leaving the market. This formation decision thus provides the underlying trading structure. For any realization, traders simply trade based on the trading rule on which they agree.

We show that, in equilibrium, traders with more volatile preferences always match with traders with less volatile preferences. 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, whereas the realization of the stable types is always zero. Note that as long as the preference shocks among extreme types are weakly 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 a result can be extended to an environment with continuous types, dynamic trading, and heterogeneous correlation.

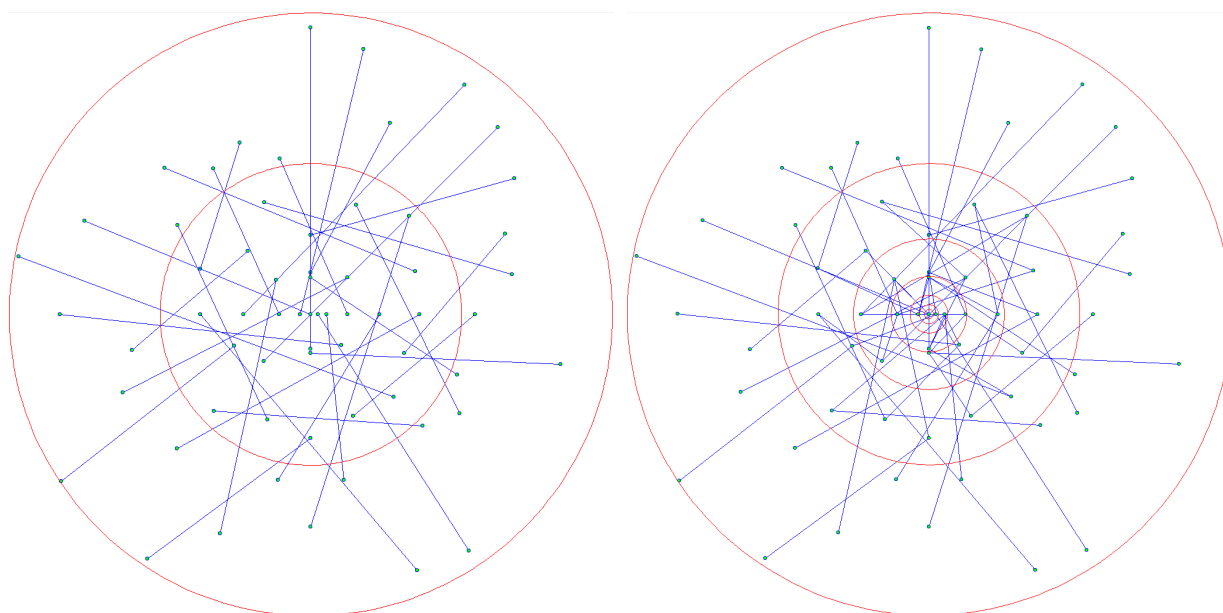
The intuition is very simple: matching involving the stable type provides insurance against the state in which extreme types have similar realizations, which implies a higher loss from misallocation. That is, the stable trader has a comparative advantage to be a market maker who always takes the opposite position of the extreme traders. Although the market maker himself might not need to trade, and although 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. Since 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. The variable  $N$  can be interpreted as potential trading opportunities in the market. As in the static model, a trader with more volatile preferences always matches with a more stable type. Moreover, it is optimal for him to leave the market afterward, because he has already reached 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 among active traders. This recursive property makes our dynamic model tractable. We can therefore analytically characterize traders' equilibrium payoff, matching decisions, and 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 farther away from the center, the more volatile the trader's preferences. The line simply represents the link between two traders. Figure 1(a)

shows the result for our one-period model: there exists a cutoff type such that traders above the cutoff type (customers) match with traders below the cutoff type (market maker).

In a dynamic model with multiple rounds, we call a trader a “dealer” if, given the realization, he trades in both directions (i.e., both buys and sells) 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 matching with dealers. This explains why traders that are far away from the center have fewer links and vice versa. Hence, traders who have more stable preferences stay in the market longer and have more trading links in equilibrium. More important, since they always match with a more volatile type in each period, they simply trade based on customers’ needs.



(a) Trading structure with one trading round ( $N = 1$ )

(b) Trading structure with multiple rounds ( $N = 7$ )

Figure 1: Market structure.

The model therefore endogenously generates a core-periphery structure. Traders who do not need to trade turn out to be the core of the network: they are the most connected and have the highest gross trade volume. Moreover, this core-periphery structure implies heterogeneity in dealers’ connectedness, as documented in Li and Schürhoff (2014)[24].<sup>3</sup> The most volatile types are purely customers, because they only trade once and leave the market right away. Traders with mid range volatility act like peripheral dealers

<sup>3</sup>Li and Schürhoff (2014)[24] find 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)[26] finds that more efficient central dealers provide intermediation services to peripheral dealers.

in the sense that they serve customers in the first period, but then they match with a more “central” dealer next period. The model also implies that central dealers gain the highest intermediation profits over time.

Since the role of market making is endogenous, we can then analyze how the intermediation profit (which maps to the 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 capture the fact that trading or building connections can be costly. On the other hand, since we allow traders to choose their trading partners based on their volatility, the correlation of traders’ preferences then simply captures the information friction. When preferences are perfectly negatively correlated, traders know with certainty who is their best counterparty. Infinite trading opportunities and perfectly negative correlations thus represent frictionless markets. We show that the profit from market making indeed increases with underlying frictions. In the limit case of a frictionless market, all traders’ payoffs converge to the payoffs implied by a competitive centralized market, that is, a zero bid-ask spread.

## Related Literature

The underlying frictions in our model are in fact implicit assumptions in the standard random search and matching model, where the counterparty arrives only at an exogenous rate (see Duffie, Garleanu and Pederson (2005)[12], Lagos and Rocheteau (2009)[22], Afonso and Lagos (2014)[4], and Hugonnier, Lester and Weill (2014)[19]). Our model gives two distinct predictions: trades are concentrated among certain traders and, the average trading rate is endogenous and heterogeneous across traders (see Rubinstein and Wollinsky (1987)[28]). Since our framework allows for heterogeneous valuation, it is closest to those of Afonso and Lagos (2014)[4] and Hugonnier, Lester and Weill (2014)[19]. Both papers show that agents with moderate valuation play the role of intermediaries in the sense that they buy and sell over time as they randomly match with other traders. Our paper is also related to Atkeson, Eisfeldt and Weill (2014)[6], 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 they have the highest gross notional trade volume.<sup>4</sup> None of these papers explicitly allows traders to choose with whom to trade. In contrast to a random search model, where all meetings are possible, our structure features clearly defined tiers among traders, and traders in the same tier will never trade with each other. The new insight here is that the tier of a trader is determined by his gain from trade. The one who has a lower need for need takes on misallocations from those in the tier who need immediacy. Hence, it is inefficient for a trader to meet with another trader in the same tier. This is a unique feature of our model, because we allow traders to choose trading partners optimally, and the price competition leads to an efficient outcome. Our

---

<sup>4</sup>Although we do not explicitly model bank size, one can interpret large banks as having a more diversified portfolio and therefore having less exposure in their preference shocks.

mechanism also gives different predictions on trading patterns and prices. In Section 4, we further compare the empirical implications between our model and random search frameworks.

This key feature also distinguishes us from models with an exogenous network structure in OTC markets (e.g., Gofman (2011)[15], Babus and Kondor (2012)[9], and Malamud and Rostek (2012) [25]), where certain traders are assumed to be more connected and therefore "customers" must trade through dealers. Here, we show that it is indeed optimal for customers to build links to dealers and become less active in the market; traders with fewer trading needs turn out to be the core of a trading network. Since our dynamic matching model endogenizes the trading structure, our paper is also related to recent literature on network formation (e.g., Babus (2012)[8], Hojman and Szeidl(2008) [17], Gale and Kariv(2007)[14], and Faboodi (2014)[13]).<sup>5</sup> Both Babus (2012)[8] and Hojman and Szeidl(2008) [17] predict a star structure in order to overcome information frictions and minimize the cost of building links. Faboodi (2014)[13] looks at the interbank lending market and considers two types of banks: banks that make risky investments over-connect, and banks who mainly provide funding end up with too few connections, as a result of bargaining frictions.<sup>6</sup> We are the first paper to explain the observed core-periphery structure that has multiple tiers. The novel prediction here is that that financial institutions that have fewer gains from trades become the core of a network endogenously. More important, we show that such a trading structure is constrained efficient subject to trading frictions in decentralized markets. Furthermore, despite the network structure, our model admits an analytical solution with rich heterogeneity, which facilitates characterizing asset price, trading volume, and trading dynamics.

The rest of the paper is organized as follows. We start with a basic model with one round trade 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. Section 5 further studies the implications of the dynamic model on properties of the trading network in OTC markets and market liquidity. Section 6 further extends the model to an infinite-horizon setup and applies the model to study the interbank lending market. Section 7 concludes the paper.

---

<sup>5</sup>See Jackson (2005)[20] for detailed review of the literature on network formation.

<sup>6</sup>Other papers that analyze bargaining frictions in the intermediation chains are, for example, Gofman (2011)[15] and Wong and Wright (2011) [29].

## 2 Basic Model: One Round of Trade

### 2.1 Setup

*The environment:* There are two periods ( $t = 0, 1$ ). There is an atomless continuum of risk-neutral traders. Agents are indexed by their preference volatility  $\sigma \in \Sigma = [\sigma_L, \sigma_H]$ , which is exogenously given and publicly observable. The function  $G(\sigma)$  denotes the measure of traders with a preference volatility below  $\sigma$ . There is one divisible asset, and all agents are endowed with  $a$  units of this asset at  $t = 0$ . Asset holdings of all traders are restricted to the  $[0, 2a]$  interval. The preference over the asset is realized at  $t = 1$ . Let  $x^i$  denote the realized preference for trader  $i$  with volatility  $\sigma$ , which is given by

$$x^i = y + \sigma x^i,$$

where  $y \geq \sigma_H$  is the expected dividend from an asset 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 an i.i.d. preference shock. Our results remain intact for any parameter  $p \in [0, 1)$ . In Section 2.4, we further impose more structure on traders' preferences in order to formalize the general environment, where correlation can be anything between 1 and  $-1$ , and it can also be heterogeneous across trading pairs. Nevertheless, since all of 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 holdings of a trader are observable to other traders, but his realized preference type is not.

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

The assumptions on preferences and the information structure are motivated by empirical works on the trading partners in OTC markets. Afonso, Kovner and Schoar (2013)[2] show that banks in the interbank lending market borrow from and lend to each other, and the fluctuations in their liquidity needs tend to be

negatively correlated.<sup>7</sup> Our model is able to capture these empirical facts with these assumptions. The idea that traders might not know who is the best counterparty is also the key implicit assumption in all random search models (see [12, 22, 4, 19]). The key difference here is that we allow traders to choose their trading partners optimally, because they can act upon the information given by the observable characteristics of their counterparties, such as preference volatility, asset holdings, 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  $S_{zz}$  represent the set of possible realizations of preferences within pair  $(z, z) = (a, a)$ . The contracts exchanged within pair  $(z, z)$  are then denoted by  $c = \{(a(z, s), t(z, s), a(z, s), t(z, s))\}$ , which specifies the amount of asset  $a(z, s)$  and payment  $t(z, s)$  received by agent  $z$  in state  $s = \{ \omega, \omega' \} \in S_{zz}$ . Define  $\mathcal{C}$  to be the set of feasible contracts,  $c \in \mathcal{C}$ , 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  $c \in \mathcal{C}$ , let  $W(z/z)$  denote the expected value for trader  $z$  who matches with  $z$ , which yields

$$W(z/z) = E_s[ a(z, s) + t(z, s)].$$

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

$$\Omega(z, z) = \max_c W(z/z) + W(z/z)$$

What matters is how the gain  $\Omega(z, z)$  is divided between the pair; therefore, we work on traders' payoffs directly instead of specifying the details of the contract. Let  $f(z, z)$  denote the measure of the pair  $(z, z)$ . Hence, if  $f(z, z) = 0$ , we say that agents  $z$  and  $z$  are not paired.

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

- 1) *Optimality of traders' matching decisions:*

$$W(z) = \max_{\tilde{z}} \Omega(z, \tilde{z}) - W(\tilde{z}) \tag{1}$$

---

<sup>7</sup>See table 3 of the paper ([2]), "determinants of existing relationships." Two key variables are among the determinants of trading relationships: The first variable is "symmetry," which measures whether banks in a relationship both borrow from and lend to each other. The sign of the variable is significantly positive. This implies that the preferences of banks in a relationship fluctuate. The second variable is "correlation of net customer funds," which measures the correlation of the volatility type for banks in a relationship. The sign of this variable is significantly negative, which means that the volatility types of banks are negatively correlated.



and for any  $f(z, z) > 0$ ,  $z \in \arg \max_{z, \tilde{z}} \Omega(z, \tilde{z}) - W(\tilde{z})$ .

2) Feasibility of the allocation function:

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

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

In the definition  $Z$  represents the set of observable characteristics of a trader. In the basic model, to highlight the role of volatility in the trader's preference type, we focus on the case in which every trader has the same asset holdings. In the baseline model,  $Z = \Sigma \cup \{\emptyset\}$ , and a trader's type is the volatility of his preference,  $\sigma$ . A null set  $\{\emptyset\}$  is meant to allow traders to be matched with no one. That is, traders are allowed not to participate in the matching game. In the basic model,  $h(z) = g(\sigma)$ .

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

Condition (1) states that, taking other traders' payoffs 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. This condition is effectively the standard pairwise stability condition.

## 2.3 Characterization

In this subsection, we focus on characterizing sorting based on the volatility type of traders. Therefore, we assume that the volatility of traders is the only observable heterogeneity,  $Z = \Sigma \cup \{\emptyset\}$ , and every trader has  $a$  units of assets. Further, we assume that the correlation between the preference of a trader and that of his counterparty is always characterized by  $p \equiv \Pr(x^i x^j < 0) = \frac{1}{2}$ . We will rationalize 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, which 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 into two subsets, divided by an endogenous cutoff type,  $\sigma^*$ . With this feature, we then characterize the equilibrium and show that the equilibrium payoff function is unique.

Lemma 1 is the key to understanding 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 among themselves. This is true as long as preference shocks are not *perfectly* negatively correlated. To see this, we first look at the pairwise surplus within each match. Clearly, any contract that maximizes the surplus must satisfy the following condition: for  $\sigma \geq \sigma^*$ ,  $a(\sigma, s) = 2a$  if  $\sigma < \sigma^*$  and  $a(\sigma, s) = 0$  if

$\succ$ . In words, the asset allocation that maximizes the joint surplus must reflect the preference of the more volatile type within the pair: the more volatile type within the pair receives the asset whenever he has a high realization and sells the asset whenever he has a low realization, regardless of the preference of the less volatile type. This also implies that the more volatile type within the pair always reaches his first-best asset allocation, whereas, the less volatile types might not. Hence, compared with the centralized competitive market, the loss of this pairwise matching comes from the fact that the less volatile type might fail to reach his optimal asset position.

Now, consider any four traders,  $\theta_4 \geq \theta_3 > \theta_2 \geq \theta_1$ . We know that two of them might not reach the first best under the pairwise matching. However,  $\theta_4$  and  $\theta_3$  have a higher need for trade, so it would be more costly if one of them failed to reach the optimal allocation. As a result, the matching outcome that maximizes the aggregate surplus is to match both of them with more stable types separately. In this way, the total loss is minimized because it is less costly if  $\theta_2$  and  $\theta_1$  fail to reach their first best. Intuitively, if we interpret that the more stable types are the ones with a more diversified portfolio, letting these traders hold the wrong asset position is less costly because they can take on more risks.

In other words, the more stable types have a comparative advantage to act as a market maker by always taking the opposite position of “customers”. Although the market maker himself might not need to trade, and even though customers can reach a higher pairwise surplus with other customers, trading through market makers minimizes the uncertainty of the preference shocks in the economy, and such matching outcomes are always efficient. On the other hand, if the information is perfect (which is the case in which preference shocks are perfectly negatively correlated), there is effectively no uncertainty in this economy. This explains why Lemma 1 holds whenever preference shocks are not *perfectly* negatively correlated.

**Lemma 1** *Let  $\theta_4 \geq \theta_3 > \theta_2 \geq \theta_1$ , for any  $p < 1$ ,*

$$\Omega(\theta_4, \theta_3) + \Omega(\theta_2, \theta_1) < \Omega(\theta_4, \theta_1) + \Omega(\theta_3, \theta_2) = \Omega(\theta_4, \theta_2) + \Omega(\theta_3, \theta_1)$$

**Proof.** The trading surplus for each state  $s$  is  $v(s) \equiv \theta - \theta/a$ . Hence, the expected surplus  $\Omega(\theta_i, \theta_j) = E_s[v(s)] + \{W^0(\theta_i) + W^0(\theta_j)\}$  has the following expression:

$$\Omega(\theta_i, \theta_j) = a[p(\theta_i + \theta_j) + (1-p)\theta - \theta/a] + W^0(\theta_i) + W^0(\theta_j) \quad (2)$$

$$= a[\max(\theta_i, \theta_j) - (1-2p)\min(\theta_i, \theta_j)] + W^0(\theta_i) + W^0(\theta_j). \quad (3)$$

where  $W^0(\theta)$  denotes the autarky value of trader  $\theta$ .

As a result,

$$\begin{aligned} & [\Omega(\theta_4, \theta_3) + \Omega(\theta_2, \theta_1)] - [\Omega(\theta_4, \theta_1) + \Omega(\theta_3, \theta_2)] \\ &= a[(\theta_4 + \theta_2) - (1-2p)(\theta_3 + \theta_1)] - a[(\theta_4 + \theta_3) - (1-2p)(\theta_1 + \theta_2)] \\ &= a[-(\theta_3 - \theta_2) - (1-2p)(\theta_3 - \theta_2)] = -2(1-p)(\theta_3 - \theta_2) < 0. \end{aligned}$$

■

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

**Proposition 1** *Any equilibrium allocation  $f$  must satisfy the following conditions: if  $f(\cdot, \cdot) > 0$  and  $f(\hat{\cdot}, \hat{\cdot}) > 0$ ,*

$$\max(\cdot, \cdot) + \max(\hat{\cdot}, \hat{\cdot}) = \cdot_4 + \cdot_3 \quad (4)$$

where  $\cdot_i$  is the  $i$ th order statistic of  $\{\cdot, \hat{\cdot}, \hat{\cdot}\}$

**Proof.** Suppose not. Now consider an equilibrium where  $f(\cdot_4, \cdot_3) > 0$  and  $f(\cdot_2, \cdot_1) > 0$ . Note that (1) can be rewritten as  $W(\cdot) + W(\cdot) \geq \Omega(\cdot, \cdot)$  for  $\forall(\cdot, \cdot)$ . Hence, we have  $W(\cdot_4) + W(\cdot_2) \geq \Omega(\cdot_4, \cdot_2)$  and  $W(\cdot_3) + W(\cdot_1) \geq \Omega(\cdot_3, \cdot_1)$ , which implies  $\Sigma W(\cdot_j) \geq \Omega(\cdot_4, \cdot_2) + \Omega(\cdot_3, \cdot_1)$ . However, since  $f(\cdot_4, \cdot_3) > 0$  and  $f(\cdot_2, \cdot_1) > 0$ , this implies that  $W(\cdot_4) + W(\cdot_3) = \Omega(\cdot_4, \cdot_3)$  and  $W(\cdot_2) + W(\cdot_1) = \Omega(\cdot_2, \cdot_1) \implies \Sigma W(\cdot_j) = \Omega(\cdot_4, \cdot_3) + \Omega(\cdot_2, \cdot_1) > \Omega(\cdot_4, \cdot_2) + \Omega(\cdot_3, \cdot_1)$ . This is a contradiction by Lemma 1. ■

**Corollary 1** *There exists  $\cdot \in [L, H]$  such that  $f(\cdot, \cdot) = 0$  for each  $(\cdot, \cdot) \in \Sigma_C \times \Sigma_C$  and  $(\cdot, \cdot) \in \Sigma_M \times \Sigma_M$ , where  $\Sigma_M = [L, \cdot]$  and  $\Sigma_C = [\cdot, H]$ .*

According to the above corollary, we can effectively solve the equilibrium as the standard assignment model with a two-sided market. For convenience, we call traders with higher volatility  $\cdot \in \Sigma_C$  “customers”, because they are the ones with a higher need for trade. And we call traders with lower volatility  $\cdot \in \Sigma_M$  “market-makers”, because they are the ones who have a lower need for trade ex ante but always provide liquidity for customers in equilibrium. These 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 of customers and market makers. In other words, how is the surplus split between customers and market makers? And what is the profit of market makers from providing such liquidity?

First, as we know from the two-sided market, complementarity determines the sorting. With Corollary 1, (2) can be conveniently rewritten as

$$\Omega(\cdot_c, \cdot_m) = a[\cdot_c + (2p - 1)\cdot_m] + W^0(\cdot_c) + W^0(\cdot_m), \quad (5)$$

where  $\cdot_c \in \Sigma_C$  and  $\cdot_m \in \Sigma_M$ . Clearly, the additive nature of the payoff implies that there is no complementarity 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. Intuitively, the loss of aggregate surplus comes from the fact that market makers might not reach their optimal allocation. Such loss is independent of which customers they match. Hence, there is no gain from any sorting between customers and market makers.

**Proposition 2** For any  $p < 1$ , a unique equilibrium payoff  $W(\cdot)$  is given by

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

where  $\cdot$  solves

$$\int_0^{\cdot} dG(\tilde{\cdot}) = \int_{\cdot}^{\cdot} dG(\tilde{\cdot}).$$

**Proof.** See the Appendix. ■

The above construction shares the standard property in the assignment model: the additional payoff gained by trader  $\cdot$  is exactly his contribution to the surplus within the match:  $W(\cdot) = \Omega(\cdot, \tilde{\cdot})$  given that he matches with  $\tilde{\cdot}$  in equilibrium. In particular, from equation (5), conditional on customer  $\cdot_c$  matching with market-maker  $\cdot_m$ , the marginal contribution of a customer is given by  $\Omega_{\cdot_c}(\cdot_c, \cdot_m) = a$ ; whereas, the marginal contribution of a dealer is represented by  $\Omega_{\cdot_m}(\cdot_c, \cdot_m) = (2p-1)a$ . This then explains the shape of the equilibrium payoff function  $W(\cdot)$ .

## 2.4 Correlation of Preferences across Traders

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

$$\cdot^i = y + x_k^i,$$

and  $x_k^i$  is given by

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

where  $\bar{X}$  and  $X_j$  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. The variable  $l_k$  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 the 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  $\forall j \neq i$ . Therefore, 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 important, the cross-group correlation is always (weakly) smaller than the within-group correlation, since  $p_1 \equiv \Pr(x_k^i x_{-k}^j < 0) = q^2 + (1 - q^2)\frac{1}{2} \geq p_0$  for  $\forall q$ . The equality holds when  $q = 0$ , which is the special case that all traders' preferences 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 preferences across groups are perfectly negatively correlated. This is, then, the environment in which information is perfect: all traders know who the best counterparty is.

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

$$\Omega(z, \tilde{z}) = a \max(\tilde{v}_k, \tilde{v}_{\tilde{k}}) - (1 - 2p_{k\tilde{k}}) \min(\tilde{v}_k, \tilde{v}_{\tilde{k}}) + W^0(\tilde{v}_k) + W^0(\tilde{v}_{\tilde{k}}),$$

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 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, because the trading surplus is higher when agents' preferences are more negatively correlated. The problem is then reduced to the one-dimensional sorting on volatility established in our basic model, in which the effective correlation in equilibrium is given by the cross-group correlation.

**Proposition 3** *Any equilibrium allocation  $f$  must satisfy Proposition 1 and  $f((k, \tilde{k}), (\tilde{k}, k)) = 0, \forall k, \tilde{k} \in \Sigma$ .*

**Proof.** See the Appendix. ■

According to Proposition 3, the equilibrium allocation can be effectively reduced to our basic model by solving  $f(k, \tilde{k})$  and setting the correlation parameter to be  $p = \max\{p_0, p_1\} = p_1$ . Since 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 as in the previous subsection, and the equilibrium payoff is then given by the following proposition.

**Proposition 4** *For any  $q \in [0, 1]$ , an unique equilibrium payoff  $W(k)$  is given by*

$$W(k) = W(\tilde{k}), \forall k \in \{A, B\},$$

where  $W(\tilde{k})$  is defined in Proposition 2 by setting  $p = p_1 = q^2 + (1 - q^2)\frac{1}{2}$ .

## 2.5 Implementation of the General Contract and Bid-Ask Spread

A general contract,  $\mathcal{C} = \{(a(z, s), t(z, s)), (a(\tilde{z}, s), t(\tilde{z}, s))\}$ , specifies the allocation of assets and the transfer of numeraire goods. In general, the numeraire goods transfer depends on the state  $s$ , bause this is a market

with spot transactions. In this subsection, we implement a general contract by a spot transaction contract, which specifies the price of the transaction for each unit of assets and total trade volume. Since we can interpret one trader in a transaction as the market maker, with  $\omega \leq 0$ , we call the bid price, denoted by  $q^b$ , if the market maker buys from his customer, with  $\omega > 0$ , and the ask price, denoted by  $q^a$ , if the market maker sells to his customer.<sup>8</sup> The equilibrium payoff can be simply implemented by market makers posting a bid price and an ask price to all customers. The bid-ask spread is proportional to the market maker's profit from trade. Denote the volatility type of a market maker to be  $\omega_m$  and his asset holdings  $a_m$ . Denote the volatility type of a customer to be  $\omega_c$  and his asset holdings  $a_c$ . Thus, the spot transaction contract between a market maker of type  $z_m = (\omega_m, a_m)$  and a customer of type  $z_c = (\omega_c, a_c)$  is

$$\begin{aligned} \{a(z_m, s), t(\omega_m, s)\} &= \{2a, -q^b[2a - a_m]\}, \text{ when } \omega_c < \omega_m, \\ \{a(z_m, s), t(\omega_m, s)\} &= \{0, q^a a_m\}, \text{ when } \omega_c > \omega_m. \end{aligned}$$

In the basic model, every trader has  $a$  units of asset (i.e.,  $a_m = a$ ). Therefore, the trade volume between a market maker and a customer is always  $a$ :

$$\begin{aligned} W(\omega_m) &= \frac{1}{2} [2aE(\omega_c > \omega_m) - q^b a] + \frac{1}{2} q^a a \\ &= ya + (2p - 1) ma + a \frac{q^a - q^b}{2}, \\ W(\omega_c) &= \frac{1}{2} q^b a + \frac{1}{2} [(y + \omega_c)2a - q^a a] \\ &= (y + \omega_c)a - a \frac{q^a - q^b}{2}. \end{aligned}$$

According to Proposition 2, the bid-ask spread is then given by  $q^a - q^b = 2(1 - p)$ , where  $q^a = y + (1 - p)$  and  $q^b = y - (1 - p)$ .

### 3 Dynamic Model: Multiple Rounds of Trade

In this section, we extend the basic model to a dynamic setting with  $N$  rounds of trade. Our basic model is essentially the special case in which all traders have only one round of trade, where we show that the stable types endogenously act like market makers in the sense that they always take the opposite position of their customers. By allowing multiple rounds of trade, our model can further study intermediation in the sense that certain traders end up buying and selling over time. Hence, the role of dealers also arises endogenously in this dynamic setting. As in the basic model, the key decision is the traders' matching decision. The only difference is that traders now choose with whom to connect for each round of trade and the number of traders with whom to connect. That is, although each trader can potentially contact  $N$  traders, the trader

---

<sup>8</sup>In general, the price could be a function of trade volume. In our current setup, it is only a function of the direction of trade.

can choose to stop matching traders at some point in time (effectively, the trader stops participating in the market). As a result, a richer trading structure emerges, where all trading links as well as the number of links for each trader are determined in equilibrium.

### 3.1 Setup and Equilibrium

The economy lasts  $N + 1$  periods ( $t = 0, 1, \dots, N$ ). To fix this idea, one can interpret our model as a intraday trading game. The parameter  $N$  therefore represents the number of trading rounds within a trading day, which captures the underlying friction that prevents traders from connecting with infinite traders. Traders receive a payoff from holding an asset at the moment they leave the market. This can be interpreted as their expected return from investment using the asset. Traders discount future payoffs using discount factor  $\delta$ ,  $\delta \in (0, 1]$ .<sup>9</sup>

In period 0, a trader chooses his trading partner for each period, from period 1 to period  $N$  based on their volatility type  $\theta$ , asset holdings, and which group the agent is in. Therefore,  $Z = \cup\{\emptyset\} \times [0, 2a] \times \{A, B\}$ . Note that the matching decision in each period can be contingent on the asset holdings in our dynamic setting. In the static model, asset holding does not play a role, because all traders have the same endowment to begin with. In the dynamic model, however, traders might have different asset positions over time, depending on their trading history. The fact that we allow for the trading decision to be contingent on asset holding implies that we assume asset positions are observable to the market. Hence, consistent with the basic model, the only uncertainty in this economy is the realized preference of traders.

**Definition 2** *An equilibrium is an allocation function,  $f_t : Z^2 \rightarrow \mathbb{R}_+$ , and equilibrium payoff  $W_t(\cdot) : Z \rightarrow \mathbb{R}_+$ , and a participation function  $\mu_t(\cdot) : Z^2 \rightarrow \{0, 1\}^2$  for  $t = 1, \dots, N$  such that the following conditions are satisfied:*

1) *Optimality of traders' matching decisions:*

$$W_t(z) = \max_{\tilde{z} \in Z} \Omega_t(z, \tilde{z}) - W_t(\tilde{z}), \quad (6)$$

and for any  $f_t(z, z) > 0$ ,  $z \in \arg \max_{\tilde{z} \in Z} \{\Omega_t(z, \tilde{z}) - W_t(\tilde{z})\}$  and  $\Omega_t(z, z) = \hat{\Omega}_t(z, \mu_t(z/z), z, \mu_t(z/z))$ .

2) *Optimality of traders' participation decisions:*

$$(\mu_t(z/z), \mu_t(z/z)) \in \arg \max_{(\mu, \mu) \in \{0, 1\}^2} \hat{\Omega}_t(z, \mu, z, \mu).$$

3) *Feasibility of the allocation function:*

$$\int_Z f_t(z, \tilde{z}) d\tilde{z} = h_t(z), \forall z \in Z, t, \quad (7)$$

---

<sup>9</sup>The discount factor  $\delta$  can also be interpreted as the probability of retaining an investment opportunity. We use this interpretation in section 5 when we use our model to study the interbank market.

where  $h_t(z)$  is the density function of  $z$  conditional on  $\pi_{t-1}(z) = 1$ .<sup>10</sup>

Equilibrium condition (6) concerns pairwise stability for any period  $t$ , where  $W_t(z)$  represents the expected value of trader  $z$ . Moreover, we will show that not all traders will remain active for  $N$  periods. Hence,  $f_t(z, \{\emptyset\}) > 0$  represents that a trader  $z$  who stays in the market but matches with no one  $\{\emptyset\}$  at period  $t$ . Traders can also choose to leave the market after they trade. Such participation decision is embedded in  $\Omega_t$  in the equilibrium definition. Formally, the sum of two traders' expected payoffs after matching,  $\Omega_t$ , depends on the asset allocation — which is given by the state contingent contract,  $\pi = \{(a(z, s), t(z, s), a(\tilde{z}, s), t(\tilde{z}, s))\}$ — as well as their optimal participation decision after they trade. Hence, given any two traders  $z = (\pi, a, i)$ ,  $\tilde{z} = (\tilde{\pi}, \tilde{a}, \tilde{i})$ , where  $\pi, \tilde{\pi} \in \mathcal{Z}$ ,  $a, \tilde{a} \in [0, 2a]$ , and  $i, \tilde{i} \in \{A, B\}$ , the maximize surplus that can be generated by this pair is then given by

$$\Omega_t(z, \tilde{z}) = \max_{(\pi, \tilde{\pi}) \in \mathcal{Z}^2} \hat{\Omega}_t(z, \pi, \tilde{z}, \tilde{\pi}),$$

where

$$\begin{aligned} \hat{\Omega}_t(z, \pi, \tilde{z}, \tilde{\pi}) = & \max_{\pi} \pi(1 - \pi) E_S[\pi a(z, s)] + E[\pi W_{t+1}(\pi, \pi, a(z, s), i)] \\ & + (1 - \pi) E[\pi a(\tilde{z}, s)] + \tilde{\pi} E_S[\tilde{\pi} W_{t+1}(\tilde{\pi}, \tilde{\pi}, a(\tilde{z}, s), i)]. \end{aligned}$$

If a trader- $z$  chooses to leave the market after the trade at period  $t$ , (i.e.,  $\pi = 0$ ), he receives the payoff from holding an asset immediately, which is given by  $E_S[\pi a(z, s)]$ . On the other hand, if he stays for another period, he then receives the continuation value next period,  $E[\pi W_{t+1}(\pi, \pi, a(z, s), i)]$ .

### 3.2 Characterization

Although the matching decision is multidimensional in our setting,  $Z = \mathcal{Z} \cup \{\emptyset\} \times [0, 2a] \times \{A, B\}$ , we know that, conditional on all other observable characteristics, (1) according to Section 2.4, group  $A$  traders will always choose to match with group  $B$  traders because their preferences are more negatively correlated; (2) because of linearity in traders payoffs, traders will always hold either 0 or  $2a$  units of assets for any period  $t > 1$ . Hence, traders with assets of  $2a$  will always choose to match with traders with no assets, because there is no gain from trade if both of them hold  $2a$  units of assets or zero assets. Given this implicit knowledge, we can then simplify the sorting problem to a one-dimensional problem in which the key variable is the volatility type. For notational convenience, we denote the sorting outcome with respect to the simplified state space,  $Z = \mathcal{Z} \cup \{\emptyset\}$ . Nevertheless, whenever we say that trader  $z$  matches with trader  $\tilde{z}$ , it is implicit that they are from different groups and have different asset holdings. That is,  $z = (\pi, 2a, A)$  and  $\tilde{z} = (\tilde{\pi}, 0, B)$ .

Before we formally construct the equilibrium, we first illustrate intuitively how it works for a two-period model (i.e.,  $N = 2$ ). The general principle is the same as in the basic model: some traders might not reach

---

<sup>10</sup>  $\pi_0(z) \equiv 1$ , for all  $z \in Z$ .



their optimal asset allocation, and the aggregate surplus is maximized when the asset is not in the wrong hands of the more volatile types (because they are the ones who have a higher need for trade). With this logic, we construct our equilibrium as follows. In the first period, just as in the basic model, all traders above the cutoff  $\theta > \theta_1$  match with traders below the cutoff type. This guarantees that all traders above the cutoff have reached their optimal asset position: if the preference shock they receive is positive, they will hold  $2a$  units of assets, and if the preference shock they receive is negative, they will hold 0 units of assets.

On the other hand, all traders below the cutoff type may have either 0 unit of assets or  $2a$  units of assets, regardless of the realization of their preference type. This implies that there is still a gain from trade for these traders. Among these traders ( $\theta < \theta_1$ ), we know that the optimal matching rule is again to guarantee that the relatively volatile ones can reach their optimal asset position. Hence, similarly, there is a cutoff  $\theta_2$  in the second period, where all active traders above the cutoff  $\theta > \theta_2$  should match with traders below the cutoff type  $\theta < \theta_2$ . One can see that, when  $N = 2$ , all traders above the cutoff type  $\theta_2$  are guaranteed to reach their efficient allocation under this matching rule. Compared with the perfectly competitive market, some assets are still in the wrong hands, but such loss is minimized because it only happens for the more stable types  $\theta < \theta_2$ .

According to the previous example, traders with more volatile preferences will be customers and quit trading earlier. In what follows, we will show that this construction can be extended to  $N$  rounds of trade. Only traders with the lowest volatility will act as market makers and stay active in all  $N$  rounds of trade. To be precise, we call traders “customers” in period  $t$  only if they are expected to reach their efficient asset allocation in period  $t$  and leave the market afterward. Similarly, we call traders “dealers” or “market makers” in period  $t$  if they always trade based on the customer’s preference and they stay for the next period and, possibly, engage in different trading directions. Note that although all traders are ex-ante the same, the role of customers, market-makers or dealers emerge endogenously in our model.

Furthermore, as shown from the previous example, the equilibrium follows a recursive structure. The matching rule at period  $t$  is characterized by a cutoff volatility type  $\theta_t$  such that all remaining traders above the cutoff type match with traders below the cutoff type. With this tractable feature, what is left to pin down is the transfer within each pair for each period. In other words, we need to solve for the transfer so that it is optimal for all traders to follow the constructed matching rule.

**Equilibrium Assignment** Given the recursive property of our dynamic equilibrium, the evolution of the distribution is straightforward. Formally, the equilibrium distribution is summarized by cutoff types  $\theta_t$  such

that  $G(\theta_t) = 2^{-t}, \forall t \in \{1, \dots, N\}$ :

$$\int_{L_t}^{\theta_t} f_t(\theta, \tilde{\theta}) d\tilde{\theta} = g(\theta_t), \quad \text{if } L_t \leq \theta_t \leq \theta_{t-1}, \quad (8)$$

$$\int_{L_t}^{\theta_t} f_t(\theta, \tilde{\theta}) d\tilde{\theta} = g(\theta_t), \quad \text{if } \theta_t < L_t \leq \theta_{t-1}, \quad (9)$$

$$\int_{L_t}^{\theta_t} f_t(\theta, \tilde{\theta}) d\tilde{\theta} + f_t(\theta_t, \{\emptyset\}) = 0, \quad \text{if } \theta_{t-1} < L_t \leq H, \quad (10)$$

where  $\theta_0 = H$ . The assignment for traders of type  $\theta \in (L_{t-1}, H]$  is 0 because they do not participate in period  $t$  trade. That is, for each period  $t$ , active traders are the ones who were dealers last period, since once traders become customers, they have reached their efficient allocation and have left the market. This is formally expressed in equation (10). Among those active traders, the more volatile types must match with the stable types, which is captured in equations (8) and (9).

**Equilibrium Payoff** According to Section 2.5, we know that this matching rule can be conveniently implemented by having market makers post bid-ask spreads. Hence, effectively, we just need to solve for the bid-ask spread in each period that implements the trading dynamics. For the last period, the transfer can be determined in a way similar to the static game. The only difference is that not all traders remain active in the last period. Nevertheless, for those who remain active, the payoff is then given by Proposition 2:

$$\begin{aligned} & [y - (1 - 2p)]a + T_N, \quad \forall \theta \leq \theta_N, \\ W_N(\theta) = & \frac{1}{2}(y - \theta)0 + \frac{1}{2}(y + \theta)2a - T_N, \quad \forall \theta_{N-1} < \theta \leq \theta_{N-1}, \\ & \frac{1}{2}(y - \theta)0 + \frac{1}{2}(y + \theta)2a, \quad \forall \theta_{N-1} < \theta, \\ & [y - (1 - 2p)]a + T_N, \quad \forall \theta \leq \theta_N, \\ = & (y + \theta)a - T_N, \quad \forall \theta_{N-1} < \theta \leq \theta_{N-1}, \\ & (y + \theta)a, \quad \forall \theta_{N-1} < \theta, \end{aligned} \quad (11)$$

where the transfer is given by

$$T_N = (1 - p) \theta_N a. \quad (12)$$

The payoff of inactive traders at period  $N$ , who are of type  $\theta > \theta_{N-1}$ , is  $(y + \theta)a$  if they make their investment at period  $N$ . So, in equilibrium, they prefer make their investment as soon as they become customers. Similarly, we can derive recursively the traders' payoff at period  $N - 1$ ,

$$\begin{aligned} & W_N(\theta) + T_{N-1}, \quad \forall \theta \leq \theta_{N-1}, \\ W_{N-1}(\theta) = & (y + \theta)a - T_{N-1}, \quad \forall \theta_{N-2} < \theta \leq \theta_{N-2}, \\ & (y + \theta)a, \quad \forall \theta_{N-2} < \theta. \end{aligned} \quad (13)$$

To pin down the transfers, we need to make sure that the marginal type is indeed indifferent between being a customer and being a dealer today. The payoff for type  $\theta_{N-1}$  to be a dealer is  $W_N(\theta_{N-1}) + T_{N-1}$ , where  $W_N(\theta_{N-1}) = y + \theta_{N-1}a - T_N$ , since  $\theta_{N-1} \geq \theta_N$ . The payoff for a type  $\theta_{N-1}$  to be a customer and leaving the market is  $y + \theta_{N-1}a - T_{N-1}$ . Therefore,

$$T_{N-1} = \frac{1}{2} (1 - \alpha) (y + \theta_{N-1}a + T_N). \quad (14)$$

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

$$\begin{aligned} W_{t+1}(\theta) + T_t, \quad \forall \theta \leq \theta_t, \\ W_t(\theta) = (y + \theta)a - T_t, \quad \forall \theta_{t-1} < \theta \leq \theta_t, \\ (y + \theta)a, \quad \forall \theta < \theta_{t-1}. \end{aligned} \quad (15)$$

Since  $\theta_t \geq \theta_{t+1}$ , the payoff for type  $\theta_t$  to be a dealer is  $W_t(\theta_t) + T_t = (y + \theta_t)a - T_{t+1} + T_t$ . The payoff for a type  $\theta_t$  to be a customer is  $(y + \theta_t)a - T_t$ . Hence,

$$T_t = \frac{1}{2} [(1 - \alpha)(y + \theta_t)a + T_{t+1}]. \quad (16)$$

For an  $N$ -period trading game, a trader's expected payoff at period 0 can be understood as the sum of his expected final asset position plus the total transfer that he has been receiving or paying over time. For a trader who stays for  $t$  periods, he will act as a market maker for  $t - 1$  periods, receiving  $\sum_{s=1}^t T_s$  from market making, and becoming a customer at period  $t$ . Once he acts as a customer, he pays for the spread  $T_t$ , reaches his efficient asset allocation  $(y + \theta)a$ , and leaves the market after period  $t$ . Hence, by construction, a trader's expected payoff has the following expression:

$$\begin{aligned} (y + \theta)a - T_1, \quad \forall \theta \in [\theta_1, \theta_H], \\ W_0(\theta) = W_1(\theta) = \sum_{s=1}^t (y + \theta_s)a + \sum_{s=1}^{t-1} T_s - T_t, \quad \forall \theta \in [\theta_t, \theta_{t-1}], t \in \{2, \dots, N\}, \\ \sum_{s=1}^N [y - (1 - 2p)\theta_s]a + \sum_{s=1}^N T_s, \quad \forall \theta \in [\theta_L, \theta_N], \end{aligned} \quad (17)$$

where

$$\begin{aligned} T_t = \frac{1-\alpha}{2} \sum_{s=t}^{N-1} 2^t \frac{1}{2^s} (y + \theta_s)a + 2^t \frac{1}{2^N} T_N, \quad \text{if } t < N, \\ T_N, \quad \text{if } t = N, \\ T_N = (1 - p) \theta_N a. \end{aligned}$$

Note that as long as traders become customers at some point, their expected final asset position is always  $(y + \theta)a$ . Hence, this explains the first part of equation (17):  $\sum_{s=1}^t (y + \theta_s)a$ . On the other hand, for traders who always act as dealers for  $N$ -periods, they expect to receive  $\sum_{s=1}^N T_s$  but might not reach their optimal asset allocation. In particular, with 50% probability, they end up with  $2a$  units of assets (which happens

when they meet a customer who has a low realization). In this case, the expected preference of the dealers is then  $E(\tilde{v}_m > \tilde{v}_c) = y - (1 - 2p)$ . Hence, the sum of these two positions gives the second part of 17.

**Proposition 5** *Assuming that  $\tilde{v}^- = \frac{y + \frac{N(2p-1)}{y+Np}}$ , in an economy with  $N$  rounds of trade, the dynamic equilibrium follows a recursive structure, where matching at period  $t$  is characterized by a cutoff volatility type,  $\tilde{v}_t$ , such that  $G(\tilde{v}_t) = \frac{1}{2^t}$ , for  $t = 1, \dots, N$ .*

- Traders participating in period- $t$  with  $\tilde{v} > \tilde{v}_t$  match with traders participating in period- $t$  trade with  $\tilde{v} \leq \tilde{v}_t$ , and they leave the market after the trade:  $\tilde{v}_t(\tilde{v}) = 0$  for  $\forall \tilde{v} > \tilde{v}_t$ .
- For all  $t = 1, \dots, N$ , traders with  $\tilde{v}$  a between  $\tilde{v}_{t+1}$  and  $\tilde{v}_t$  stop participating the market after period  $t + 1$ . That is,  $\tilde{v}_j(\tilde{v}) = 1$  for  $j \leq t$  and  $\tilde{v}_j(\tilde{v}) = 0$  for  $j \geq t + 1$ .<sup>11</sup>
- The equilibrium distribution is characterized by equations (8), (9), and (10). The equilibrium payoff of traders is determined by equations (11) ~ (17).

**Proof.** See the Appendix. ■

To understand why the constructed equilibrium indeed satisfies the trader's optimality condition, we must show that, at any point of time  $t$ , there is no profitable deviation for a customer  $\tilde{v} > \tilde{v}_t$  to match with another customer, or for a dealer  $\tilde{v} < \tilde{v}_t$  to match with another dealer. Note that any bilateral matching can at most guarantee that one trader has reached his first-best allocation. A trader who acts as a market maker in period  $t$  takes on the possible misallocation within the pair; on the other hand, being a customer guarantees that he can reach the optimal allocation faster. Hence, what determines the equilibrium is effectively who is more willing to take on the misallocation and who is more willing to pay for immediacy. Such an incentive can be seen clearly in the following claim, which establishes that equation (17) is the solution to the following optimization problem.

**Claim 1** *Equilibrium payoff function (17) solves the following:*

$$W_0(\tilde{v}) = \max_{\tilde{v}} [v(t, \tilde{v}) - \Gamma(t)],$$

where

$$v(t, \tilde{v}) = \begin{cases} \tilde{v}^t (y + \tilde{v}) a, & \text{if } t \leq N, \\ \tilde{v}^N [y - (1 - 2p)] a, & \text{if } t = N + 1, \end{cases} \quad (18)$$

and

$$\Gamma(t) = \begin{cases} \tilde{v}^t T_t - \prod_{s=1}^{t-1} \tilde{v}^s T_s, & \text{if } t \leq N, \\ - \prod_{s=1}^N \tilde{v}^s T_s, & \text{if } t = N + 1. \end{cases} \quad (19)$$

---

<sup>11</sup>Define  $\tilde{v}_0^* = L$  and  $\tilde{v}_{N+1}^* = H$ .

In other words, the decision of being a customer or a dealer at each period  $t$  is mathematically analogous to the incentive compatibility conditions. Note that  $t = N$  and  $t = N + 1$  represent the payoff of customers and market makers at period  $N$ . Traders are effectively facing a trade-off between reaching their first-best allocation earlier and being charged for a higher transfer:  $v(t, \theta)$  is weakly decreasing in  $t$ , whereas  $\Gamma(t)$  is weakly decreasing in  $t$ .<sup>12</sup> With this interpretation, one can see that the willingness to pay for immediacy, that is, a higher  $\theta$  (and a lower  $t$ ), is increasing in  $t$ . That is, formally,  $v(t, \theta) - v(t, \theta')$  is strictly increasing in  $\theta$  for all  $t < t'$ .

Hence, in period  $t$ , according to Eq. (16), we know that the marginal type  $\theta_t$  is indifferent between being a customer and acting as a market maker. It is therefore optimal for all traders of type  $\theta > \theta_t$  to act as customers:  $W(t, \theta) = W(t + 1, \theta) \Rightarrow W(t, \theta) \geq W(t + 1, \theta)$  for all  $\theta > \theta_t$ .

**Remark** Note that equality is only possible when  $\beta = 1$ . Because there is no discounting, what matters is only who takes on the misallocation at the end of the trading game. In that case, the total payment for traders with type  $\theta \in [\theta_N, \theta_H]$  should be the same. Indeed, Eq. (16) is then reduced to  $T_{t-1} = \frac{1}{2}T_t$ , and therefore  $\Gamma(t) = -2^{-N+1}(1-p)\theta_N a$  for  $\forall t$ . This implies that all traders of  $\theta \in [\theta_N, \theta_H]$  become customers in the end, and they are also indifferent regarding when to leave the market:  $\theta(\theta) = \arg \max_t [v(t, \theta) - \Gamma(t)] = \{t | 1 \leq t < N\}$ ,  $\forall \theta \in [\theta_N, \theta_H]$ . In this case, the order of leaving the market first does not matter for all traders of type  $\theta \in [\theta_N, \theta_H]$ . That is, one can arbitrarily assign the order in which customers exit as long as the market clearing condition is satisfied. Hence, in this special case, although the prediction of the market structure remains the same, the number of links for each type  $\theta \in [\theta_N, \theta_H]$  is indeterminate.

Observe that, as in our static game, the marginal payoff of a trader with type  $\theta$  is given by the marginal contribution to the trading surplus:

$$\frac{d}{d\theta} W_0(\theta) = \begin{cases} \theta a, & \forall \theta \in [\theta_N, \theta_H], \\ \theta(2p - 1), & \forall \theta \in [\theta_L, \theta_N]. \end{cases}$$

## 4 Efficiency

In this section, we first compare the equilibrium allocations to the frictionless benchmark. Then, we establish that the laissez-faire equilibrium is constrained efficient subject to the underlying frictions.

### 4.1 The Frictionless Benchmark

Trading in decentralized markets has the following distinct features: (1) trades are bilateral and (2) traders do not know where their best counterparties are. The combination of these two features generate the underlying

<sup>12</sup>According to Eq. (16):  $\Gamma(t + 1) - \Gamma(t) = -\theta \{T_{t+1} - 2T_t\} = -\theta(1 - \beta)(c + \theta^*)a < 0$

frictions. The first feature is a stark contrast to a centralized exchange, where all traders face one market clearing price and trade is multilateral. In that case, there is no need to identify the “right” counterparty, and all traders will trade at the market clearing price. On the other hand, bilateral matching can also achieve efficient allocation immediately as long as the realized preferences of all investors are observable to market participants.<sup>13</sup> That is, the matching decision can be based on the realized preference. In that case, a trader with high realization ( $x^i = 1$ ) will match with a trader with a lower valuation ( $x^i = -1$ ). Due to the price competition, all buyers will pay the same price regardless which seller they match with. Hence, in either case, the first-best allocation is achieved: traders with high realization ( $x^i = 1$ ) end up with  $2a$  units of asset, and traders with low realization ( $x^i = -1$ ) sell their assets.

**Claim 2** *In the frictionless benchmark, in which trade either takes place in a centralized exchange market, in a bilateral market where traders’ preferences are perfectly observable, traders’ payoff is given by*

$$W_{FB}(x) = (y + x)a,$$

*and the aggregate surplus is then given by  $\Pi^{FB} = \int W_{FB}(x) dG(x)$ .*

## 4.2 The Frictionless Limit

Compared to the frictionless benchmark, trading frictions in our model are captured in the following ways, governed by three parameters:

- (1) The fact that the realized “preference” of a trader is unobservable captures the information friction. Hence, all matching decisions can only be conditional on  $x$ . As long as preferences between two traders are not perfectly negatively correlated, (i.e.  $p < 1$ ), it is possible that any bilateral contract will fail to reach the efficient allocation. The parameter  $p$  then captures the degree of information frictions, and the perfect information case is then represented by  $p = 1$ .
- (2) The fact that traders can only contact finite trading partners (i.e.,  $N$ ) captures the constraint due to the bilateral trade, and it is not feasible for a trader to contact infinite number of traders.
- (3) Lastly, delay in reaching efficient allocation can be costly. This is captured by the discount factor  $\delta$ .

To compare our results with the frictionless benchmark (where traders trade instantly at the market price), we now shut down the cost of delay by setting  $\delta = 1$ . The total profit for those “central” dealers is then given by  $1 - 2^{-N} - 2(1-p) \frac{1}{N}a$ , which converges to zero when  $N \rightarrow \infty$  or  $p \rightarrow 1$ . These two limit cases respectively represent the frictionless environment where traders can either match with infinite counterparties or their

---

<sup>13</sup>If not all traders hold  $a$  units of assets ex ante, a bilateral market with traders of observable preference types is not as efficient as a centralized market. This is because clearing is multilateral in a centralized market.

preferences are perfectly negatively correlated (no information friction). The payoff of any trader then converges to what he would have gained in a competitive market.

**Proposition 6** *When  $\beta = 1$ , traders' payoffs as  $N \rightarrow \infty$  or  $p \rightarrow 1$ , the equilibrium asset allocation and equilibrium payoffs converge to those in the frictionless benchmark.*

### 4.3 Constrained Efficient Allocation

Due to the underlying frictions, the equilibrium cannot implement the first-best allocation for any finite  $N$  and any  $p < 1$ : only the asset allocation to traders with volatility type above  $\bar{\nu}$  is efficient, while the asset allocation to traders with volatility type below  $\bar{\nu}$  is random, which implies misallocation compared to the benchmark. In this section, we ask whether the equilibrium can implement the second-best allocation, that is, the allocation that maximizes total surplus subject to constraints imposed by bilateral matching and the information friction. In other words, the planner chooses for each period matching rule  $f_t$  and market participation rule  $\mu_t$  conditional on observable information so as to maximize the total welfare.

Clearly, the trading surplus generated by each pair can at most guarantee that one of the two traders reaches his first best allocation; hence, the planner's problem can be reduced to choose (1) which trader to reach first best allocation in each period; and (2) whether the trader should remain active. With this interpretation, the planner's problem can then be written as:

$$\Pi \equiv \max_{\{f_t, \mu_t\}_{t \in \Sigma}} \sum_{t=1}^N \int_{\nu} (1 - \mu_t(\nu)) [\mu_t(\nu)(y + \nu) + (1 - \mu_t(\nu))(y + (2p - 1)\nu)] g_t(\nu) d\nu$$

such that

$$\mu_t(\nu) - \mu_{t-1}(\nu) = 1, \forall \nu \in \Sigma \text{ with } \nu \leq \bar{\nu} \quad ; \quad \mu_t(\nu) = 0, \forall \nu \in \Sigma \text{ with } \nu > \bar{\nu},$$

<sup>14</sup>and for all  $\nu \in \Sigma$ ,

$$\mu(\{s : \mu(s) = 1, s \leq \bar{\nu}\}) + \mu(\{s : \mu(s) = 0, s \leq \bar{\nu}\}) = G_t(\bar{\nu}),$$

$$G_{t+1}(\bar{\nu}) + \int_{\bar{\nu}}^{\infty} [1 - \mu(s)] dG_t(s) = G_t(\bar{\nu}).$$

The function  $\mu_t(\nu)$  indicates whether a trader of type  $\nu$  receives first best allocation at period  $t$ . And  $\mu_t(\nu)$  indicates whether a trader of type  $\nu$  still participates in period  $t$  matching and trading. When  $\mu_t(\nu) = 1$ , a trader of type  $\nu$  receives first best allocation, from which the planner receives welfare  $\mu_t(\nu)(y + \nu)$  if he decides to let the trader leave the market at period  $t$ . When  $\mu_t(\nu) = 0$ , a trader of type  $\nu$  takes on the opposite position of a trader who receives first best allocation. His contribution to social welfare, if the planner decides to let him leave the market at period  $t$ , is  $(1 - \mu_t(\nu))(y + (2p - 1)\nu)$ . The constraint guarantees that the measure of

---

<sup>14</sup>  $\mu_0(\nu) = 0$ , for all  $\nu \in \Sigma$ .

traders who start receiving first best allocations at period  $t$ ,  $\mu(\{i : \pi_t(i) - \pi_{t-1}(i) = 1, \forall i \in \mathcal{I}\})$ , cannot be greater than the measure of traders who receive random asset allocation,  $\mu(\{i : \pi_t(i) = 0, \forall i \in \mathcal{I}\})$ .

The following proposition characterizes the constrained efficient allocation.

**Proposition 7** *Assume that  $\gamma \geq \gamma^* = \frac{\gamma^+ - \gamma^-}{\gamma^+ - \gamma^-}$ , the constrained efficient allocation is as follows,*

(i) *Traders of type  $\gamma \in (\gamma^-, \gamma^*]$ , for all  $t \in \{1, \dots, N\}$ , receive first best allocation at period  $t$  and participate in the trading game until period  $t$ .*

(ii) *Traders of type  $\gamma \in [\gamma^*, \gamma^+]$  face asset misallocation and participate in all rounds of the trading game.*

**Corollary 2** *The equilibrium allocation of the OTC market is constrained efficient.*

The constrained efficient allocation is easy to understand. Due to the underlying frictions, bilateral trade and information frictions, misallocation of assets is unavoidable. Hence, the constrained efficient allocation simply minimizes the overall misallocation. Since it is less costly for the stable types to take on the misallocation, it is efficient to have the more stable types to match with the more volatile types. By doing so, the more volatile types are then guaranteed to reach their efficient allocations. The measure of traders who can reach their efficient allocations in each period are constrained by bilateral matching. In other words, at most half of active traders with misallocated assets can reach efficient allocations, at the cost of having the other half with more stable types to undertake the misallocation. This explains the results in Proposition 7. Traders with larger gains from trade reach their efficient allocations earlier, and the most stable types stay until the end and face asset misallocations. This is exactly the allocation from the laissez-faire equilibrium constructed in Section 3.

## 5 Implications

In this section, we study the implications of the dynamic model on the properties of the endogenous network structure, and market liquidity, measured by trade volume, bid-ask spread, and the total number of trading rounds. Then we compare the equilibrium of the decentralized market with the equilibrium of a competitive market operated by a fictitious auctioneer.

### 5.1 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  $\pi_t : Z \rightarrow [0, 1]$ .  $\pi_t(z) = 0$  if the trader does not participate in



period- $t$  trade.

$$l_t(\theta) = \begin{cases} 0, & \forall \theta \in [t-1, H], \\ 1, & \forall \theta \in [L, t-1]. \end{cases}$$

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

$$l(\theta) = \begin{cases} 1, & \forall \theta \in [L_1, H], \\ t+1, & \forall \theta \in [L_{t+1}, L_t], t \in \{1, \dots, N-1\}, \\ N, & \forall \theta \in [L, L_N], \end{cases}$$

which implies that more efficient market makers, those with smaller  $\theta$ , 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 fewer links in equilibrium. This is consistent with findings in Neklyudov (2013)[26]. 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 illustrated the network structure by comparing the network when  $N = 1$  and when  $N = 7$ . We locate traders of lower volatility type closer to the center of the graph.

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

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

In Figure 2, we illustrate the distribution of trading links with the case  $N = 10$ .

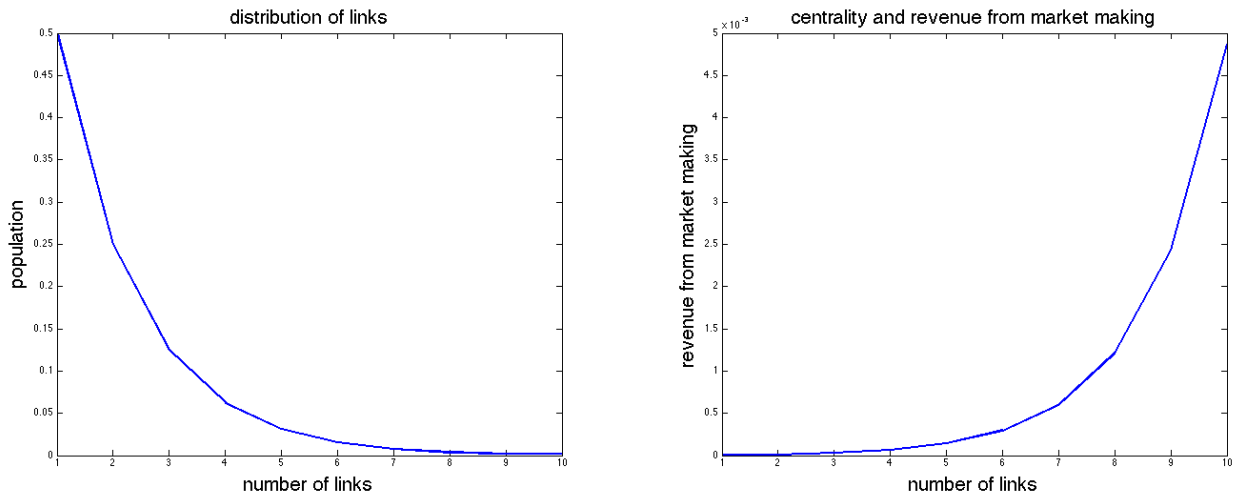


Figure 2: The distribution of trading links, revenue from market making ( $N = 10$ ).

Given the distribution of links, 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 characterize the *sparsity* of network by  $\rho(N) \equiv \frac{\bar{l}(N)}{N}$ .

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

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

## 5.2 Trade Volume

The total expected trade volume,  $EVol(\cdot)$ , conditional on type  $\theta$ , is

$$EVol(\theta) = \begin{cases} a, & \forall \theta \in [c_1, c_H], \\ a(t-1), & \forall \theta \in [c_t, c_{t-1}], t \in \{2, \dots, N\}, \\ aN, & \forall \theta \in [c_L, c_N]. \end{cases}$$

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  $c$  have 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.

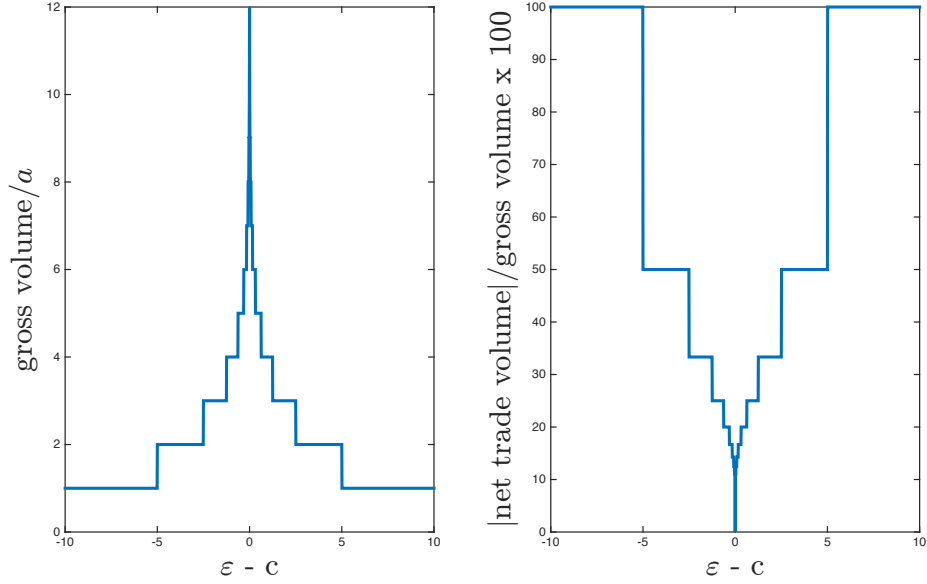


Figure 3: Trade volume across the preference type of traders. ( $N = 10$ )

### 5.3 Revenue from Market Making and Bid-Ask Spread

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

$$R(\cdot) = \begin{cases} 0, & \forall \in [1, H], \\ \sum_{s=1}^t T_s, & \forall \in [t, t-1], t \in \{2, \dots, N\}, \\ \sum_{s=1}^N T_s, & \forall \in [L, N]. \end{cases}$$

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.

There are two factors driving bid-ask spread: (1) value of immediacy and (2) benefit from avoiding future payment. When a trader pays for the spread at  $t$ , reaches efficient allocation, and leaves the market, he benefits from the immediacy of the service and also benefits from avoiding future payment.  $T_t$ , the transfer

to market makers at  $t$ , reflects the two benefits for the marginal customer with type  $\theta_t$ .

$$T_t = \underbrace{(1 - \alpha)(y + \theta_t)a}_{\text{benefit from immediacy}} + \underbrace{\frac{T_{t+1} - T_t}{2}}_{\text{benefit from avoiding future payment}}, \quad (20)$$

$$T_t = \frac{1 - \alpha^{N-t}}{2} (y + \theta_t)a + \frac{\alpha^{N-t}}{2} T_N, \forall t < N, \quad (21)$$

$$q_t^a - q_t^b = 2\frac{T_t}{a}, \quad (22)$$

$$q_t^a = y + \frac{T_t}{a}, \quad (23)$$

$$q_t^b = y - \frac{T_t}{a}. \quad (24)$$

These two factors have different implications on the time series of bid-ask spread during the trading game. If the benefit from immediacy dominates, bid-ask spread tends to decrease over time, as traders of lower volatility type value less immediacy. If benefit from avoiding future payment dominates, bid-ask spread tends to increase. For example, if  $\alpha = 1$ ,  $T_{t+1} = 2T_t$ . Figure 4 shows that the relative importance of these two factors varies when the volatility of the economy shifts. When the economy becomes less volatile,  $\alpha$  drops. As  $T_N$  is proportional to  $\alpha^{N-t}$ , the decrease reduces the relative importance of the factor (2). Afonso and Lagos (2012) [3] document the bid-ask spread during a trading day in the fed funds market. The pattern of the time series shifted after the financial crisis in 2008. Before 2008, the spread did not decrease over time and might increase at the end of a day. Since 2008, the spread has been lower in general and has shown a downward sloping trend during a trading day. As banks have held excessive reserves since 2008, their volatility types have been smaller (i.e., fewer trading needs for the interbank lending market). This shift in the pattern of the spread is therefore consistent with our model.<sup>15</sup>

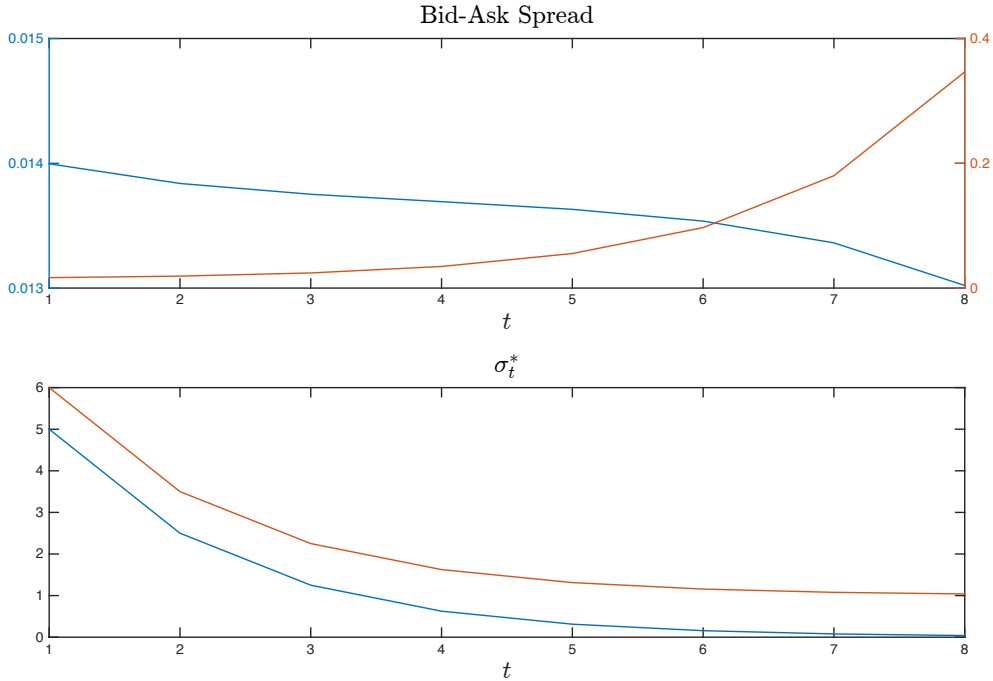
If the bid-ask spread is increasing in  $t$ , it means it is more costly to trade with more efficient dealers. This result is then consistent with findings in Li and Schürhoff (2014) [24]. Because our paper identifies two factors that drive the bid-ask spread, we provide an explanation to why we might observe different empirically patterns depending on the underlying distribution of trading needs. As an illustration, Figure 4 shows the time series of bid-ask spreads with two distributions of volatility. When the economy becomes less volatile in the sense that the distribution function  $G(\cdot)$  shifts to  $G(\cdot - \Delta)$  with  $\Delta > 0$ , the bid-ask spreads shift from an upward sloping curve to a downward sloping curve.

## 5.4 Comparison to Random Search Models

In random search frameworks, trading friction is modeled as an exogenous meeting rate (Duffie, et al. (2005) [12]), which captures the fact that it takes time to find the “right” counterparty. Based on this, recently works

<sup>15</sup>The existing literature with random matching may not be able to match this shift in pattern. See, for example,[4].

Figure 4: Time Series of Bid-Ask Spread During a Trading Day



by Afonso and Lagos (2014)[4], Hugonnier, Lester, and Weill (2014)[19] further allow for richer heterogeneity, where the valuation of a counterparty is drawn from a distribution. In their environment, traders with moderate valuation act as intermediaries as they are more likely to buy and sell given the distribution they face. In our framework, all meetings are directional. In fact, all traders in those frameworks would not trade with “intermediaries” if they could choose otherwise. Despite our mechanisms being very different, several predictions are similar here: (1) misallocation as well as trading volume are concentrated in traders with moderate valuation, and (2) allocation converges to the efficient outcome in the frictionless limit. However, our framework has several different implications regarding trading structure and prices.

First, our equilibrium structure has a defined tiering, in the sense that banks in the same tier will never trade with each other, whereas, all meetings are possible in a random search framework. The tier of a trader is determined by his gain from trade and hence his willingness to wait. Traders who are more willing to wait take on misallocations from traders in other tiers who need immediacy. Hence, it is inefficient for a trader to meet with another trader in the same tier, and that is why it never happens in our environment. This is a unique feature of our model because traders choose trading partners optimally, and price competition leads to an efficient outcome.

This key feature also leads to different price implications. In Hugonnier, Lester, and Weill (2014)[19], trading price within a pair is given by a weighted value of buyers’ and sellers’ reservation value, and such

weight is given by an exogenous bargaining power. A buyer with high valuation then pays a higher price on average. This, however, is not necessarily true in our model: buyers with higher valuation are the customers, who paid the spread in the earlier period. In fact, when  $\beta = 1$ , they pay a lower asking price. On the other hand, a buyer with slightly lower valuation (the peripheral dealer), pays a higher asking price when he leaves the market but receives spreads from other customers.

## 6 Application: the Interbank Lending Market

In this section, we apply our model to understand the systemic risk and how market liquidity affects default in the interbank lending markets, such as the Fed Funds market and the Triparty Repo market. We use the number of trading rounds,  $N$ , to study the effect of market liquidity on default risk and the location of default risk in the endogenous financial network, as the number of trading rounds represents the trading capacity of the market, which from the previous section is linked to measures of market liquidity. We study two types of default separately, default because of liquidity constraint, and strategic default induced by limited commitment.

### Interpretation of the Model as Interbank Lending

Assets being traded provide the funds a bank needs for investment. All investment pays at the end of the trading game.  $1 - \beta$  is the probability a bank may lose its investment opportunity if the bank does not invest today. A bank's preference over the asset,  $\beta$ , therefore represents the expected return on investment,<sup>16</sup>  $\beta = E\tilde{R}_i$ , where  $\tilde{R}_i$  is the realized return. The return,  $\tilde{R}_i$ , could be subject to an idiosyncratic shock.<sup>17</sup> This environment allows us to study the model implication on concentration of default risk. The contract in our model specifies both asset allocation and transfer between banks. The contract is a lending contract if the transfer is delayed. Assume that the repayment of the lending contract takes place at the end of the trading game.<sup>18</sup>

---

<sup>16</sup>Alternatively, the valuation could represent the valuation of depositors with liquidity need. To fix ideas, we focus on the interpretation that  $\beta$  is the expected return on investment.

<sup>17</sup>Volatility of the return on investment can be linked to a bank's ability to diversify risks, such as liquidity risk or idiosyncratic risk of return on assets. Commercial banks participating in the Fed Funds market can diversify the risk of liquidity demand of their customers if they have more branches. So, the volatility in our model represents the size of banks, when we apply the model to the Fed Funds market. In the repo market, an investment bank can diversify the return on its investment by diversifying its portfolio. But some banks may specialize in investing in certain assets. So, when we apply the model to the repo market, the volatility represents the ability to diversify their portfolios.

<sup>18</sup>If we think of the model in terms of trading dynamics within a trading day, this is consistent with the practice in the Fed Funds market and the Triparty Repo market, where banks borrow and lend to each other to manage their liquidity positions.

## 6.1 Interdependence and Financial Stability

In this section, we analyze how distress at a single bank may induce a cascade of defaults throughout the financial system. As banks' profit from investment, return from market making and leverage are correlated with their position in the network, banks' vulnerability to investment shocks and to the default risk of borrowers are determined endogenously. It therefore contributes to a growing literature on financial contagion [5][1], credit chains [21], and financial stability [16], where the network structure is taken as given.

Our analysis of financial contagion follows three steps: First, the proposition below establishes the numbers of banks that are "physically" connected to each other, which gives an upper bound of the maximum contagion. Second, based on banks' balance sheets in equilibrium, we characterize their vulnerability to investment shocks and to counterparty risk. Lastly, with these two vulnerability measures, we study the extent of contagion.

**Proposition 8** *Given the number of trading rounds,  $N$ , at most  $2^N$  banks are connected to each other.*

Clearly, trading capacity  $N$  improves the efficiency of asset allocation. As the above proposition shows that interconnectedness also increases with trading capacity, the model implies a tradeoff between efficiency of asset allocation in normal times and financial stability in crisis times. To understand the cost of such interdependence, we now study exactly how interconnectedness affects financial stability.

As a "market maker" repeatedly receives assets from lending banks and delivers assets to borrowing banks in the economy, loans and liability of the bank's balance sheet build up as a byproduct of the market-making activity. As the leverage ratio of a market maker builds up, he becomes vulnerable to risks in the economy. First, his exposure to counterparty default risk is high because he lends to many borrowers. Second, he himself may have incentive to default, because he borrows from many lenders.

We define the leverage ration as the debt holding of a bank over the expected asset return at the very beginning of the game, *ca.*

**Proposition 9** *The maximum leverage ratio in the market is increasing in market liquidity  $N$ . The maximum leverage is taken on by traders  $L < \leq N-1$ , i.e, traders in the "core."*<sup>19</sup>

Not only the network structure matters, but also the properties of banks that tend to play a certain role in the network. The endogenous core-periphery network structure has the following properties:

- (i) Core market makers make the most profit from market making but receive the least return on their own investment. They also build up a large balance sheet because of their market making, taking on high leverage.

---

<sup>19</sup>According to the empirical literature, traders in the core are those with the largest number of links. In our model, this includes core market makers,  $\leq N^*$ , and periphery market makers who trade with core market makers at the last round,  $N^* < \leq N_{-1}^*$ .

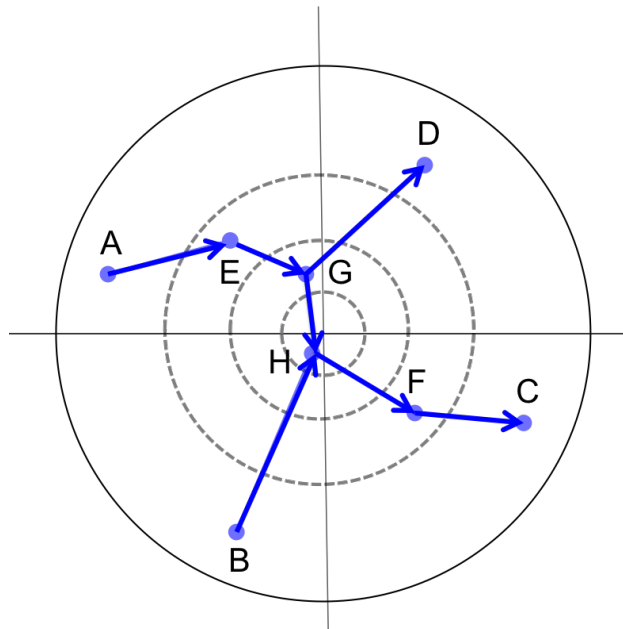
- (ii) Borrowing customers are charged the lowest spread and receive the highest return on their own investment.
- (iii) Compared with core market makers, periphery market makers make less profit from market making, but receive higher return on their own investment. Compared with customers, they are charged a higher spread when they stop making market and trade with core market makers.

Below, we use a simple example to illustrate how this endogenous network structure affects financial stability and to fully characterize the possible contagion under different parameters.

### 6.1.1 Illustrative Example: Eight Interconnected Banks when $N = 3$ .

Consider a network formed among eight banks in an economy with three rounds of trade ( $N = 3$ ). Figure 5 shows the network structure. The location of banks  $A$  to  $H$  in the network and asset flow through the network are noted in the figure. Table 1 explains the volatility type, realized expected return, and the balance sheet of each bank in the eight-bank network. The volatility type of banks,  $\sigma_i$ , follows an order statistics:  $\sigma_4 > \sigma_3 > \sigma_2 > \sigma_1$ . And the volatility types satisfy:  $\sigma_4 > \sigma_1 > \sigma_3$ ,  $\sigma_3 > \sigma_2 > \sigma_2$ ,  $\sigma_2 > \sigma_3 > \sigma_1$ . So bank  $A, B, C$  and  $D$  are customer banks, since  $\sigma_4 > \sigma_1$ . Banks  $E, F$ , and  $G$  are periphery market makers, which make the market for one or two rounds and eventually become customers of more central market makers. Bank  $H$  is the core market maker, with  $\sigma_3 > \sigma_1$ . Other notations are explained below the figure.

Figure 5: An Endogenous Network with  $N = 3$ .



Note: the direction of the links refers to the flow of funds during the day.



Table 1: Balance Sheet of Banks in the 8-Bank Network

Bank ID	Volatility Type	Expected Return	Lending	Investment Return	Borrowing
A	4	$y - 4$	$\tilde{A}q_1^b a$	0	0
B	4	$y - 4$	$\tilde{B}q_1^b a$	0	0
C	4	$y + 4$	0	$\tilde{C}(y + 4)2a$	$q_1^a a$
D	4	$y + 4$	0	$\tilde{D}(y + 4)2a$	$q_1^a a$
E	3	$y - 3$	$\tilde{E}q_2^b 2a$	0	$q_1^b a$
F	3	$y + 3$	$\tilde{F}q_1^a a$	$\tilde{F}(y + 3)2a$	$q_2^a 2a$
G	2	$y - 2$	$\tilde{G} q_1^a a + q_3^b 2a$	0	$q_2^b 2a$
H	1	$y \pm 1$	$\tilde{H}q_2^a 2a$	$\tilde{D}(y \pm 1)2a$	$q_3^b 2a + q_1^b a$

Note:  $\tilde{\delta}_i$  represents the realized repayment of loans of bank  $i$ , relative to the repayment without default or delay in debt repayment.  $\tilde{\phi}_i$  represents the ex post realized return of investment relative to the expected return. The expected return of bank  $H$  could be either  $c + \sigma_4$  or  $c - \sigma_4$ . The bid and ask prices,  $q_t^b$  and  $q_t^a$  follows Eq. (23) and (24), with  $N = 3$ .

We first look at the vulnerability of a bank to loan default and investment risk. Let  $\tilde{\delta}_i$  denote the percentage decline of total repayment to bank  $i$ , which depends on the number of defaulting borrowers and the liquidation value of default loans. Hence, if any borrowers from bank  $i$  default,  $\tilde{\delta}_i < 1$ . On the other hand, let  $\tilde{\phi}_i$  represents percentage change of bank  $i$ 's investment return; hence, whenever there is negative shock to bank  $i$ 's investment,  $\tilde{\phi}_i < 1$ . Therefore,  $\tilde{\delta}_i$  and  $\tilde{\phi}_i$  represent fluctuation in the value of bank  $i$ 's assets. When the asset value is lower than the required debt repayment, there is not enough funding to repay the debt and bank  $i$  is forced to default. How much loss a bank can take depends on their other asset positions. Hence, the thresholds for default are characterized by  $\bar{\delta}_i$  and  $\bar{\phi}_i$ .

$$\bar{\delta}_i = \frac{D_i - R_i A_i}{L_i},$$

$$\bar{\phi}_i = \frac{D_i - L_i}{R_i A_i},$$

where  $L_i$  denotes the repayment to bank  $i$ 's loan if there is no default and  $D_i$  denotes bank  $i$ 's debt obligation,  $A_i$  bank  $i$ 's investment, and  $R_i$  bank  $i$ 's investment return. Table 2 summarizes the default threshold. A higher threshold means that the bank is more vulnerable to risk. In this sense, threshold  $\bar{\phi}_i$  represents bank  $i$ 's vulnerability to investment risk. Threshold  $\bar{\delta}_i$  represents bank  $i$ 's vulnerability to default by borrowers.

**Vulnerability to Investment Risk.** Only banks making an investment in equilibrium are subject to the investment risk. They are banks  $C$ ,  $D$ ,  $F$ , and  $H$ . A bank with low  $\bar{\phi}_i$  is less vulnerable to the risk.

Table 2: Vulnerability of Banks in the 8-Bank Network

Bank ID	Volatility Type	Expected Return	$\bar{r}_i$	$\bar{r}_i$
A	4	$y - 4$	0	n.a.
B	4	$y - 4$	0	n.a.
C	4	$y + 4$	n.a.	$\frac{y+3/8}{2(y+4)}$
D	4	$y + 4$	n.a.	$\frac{y+3/8}{2(y+4)}$
E	3	$y - 3$	$\frac{y-3/8}{2y-3/2}$	n.a.
F	3	$y + 3$	0	$\frac{y+3/8}{2(y+3)}$
G	2	$y - 2$	$\frac{2y-3/2}{3y-7/8}$	n.a.
H	1	$y \pm 1$	$\frac{y-7/8}{2y+3/2}$	$\frac{y-13/8}{2(y\pm 1)}$

Banks  $C$  and  $D$  are customer banks. Bank  $F$  is a periphery market maker. And bank  $H$  is a core market maker. From Table 2, we have the following observations:

- (i) Customers are less vulnerable to investment risk than the periphery market maker, as  $\bar{r}_C = \bar{r}_D < \bar{r}_F$ . This is because the return on investment for customer banks,  $y + 4$ , is higher than that of the periphery market maker,  $y + 3$ . Another reason is that the borrowing cost of the periphery market maker,  $q_2^a$ , is higher than the borrowing cost of the customer bank,  $q_1^a$ .
- (ii) The core market maker may or may not be more vulnerable to investment risk than the periphery market maker. The relative vulnerability is affected by two determinants. First, the core market maker earns more from market making,  $q_2^a 2a - q_3^b 2a + q_1^b a > q_1^a a - q_2^a 2a$ . On the other hand, the core market maker earns less from investment,  $(y \pm 4)2a < (y + 3)2a$ .

**Vulnerability to the Default by Borrowers.** The non-monotonicity becomes more phenomenal when we look at the vulnerability of a bank to default by its borrowers. For example, bank  $F$  is not vulnerable to borrowers' default,  $\bar{r}_F = 0$ . This is because its own return on investment,  $y + 3$ , is higher than the borrowing cost,  $q_2^a$ . This measure is closely linked to financial contagion. We will use this measure to characterize financial contagion in the next part.

The take-away is that the relationship between network centrality and vulnerability to investment risk may be non-monotonic. Two factors affect vulnerability: return on own investment and net benefit from market making or trading. The two factors vary endogenously with the centrality of a bank, resulting in the non-monotonicity. The non-monotonicity highlights the importance of the endogenous network, which is a selection mechanism that correlates the volatility type of a bank and its position in the network.

Table 3: Contagion through the 8 Interconnected Banks

Parameter Value of $y$	Default Chain(s)
$0, \frac{1}{8} \text{ }_3$	$D - G$
$\frac{1}{8} \text{ }_3, \frac{5}{8} \text{ }_3$	$D - G - E$
$\frac{5}{8} \text{ }_3, \frac{7}{8} \text{ }_3$	$D - G - E, H - G - E$
$\frac{7}{8} \text{ }_3, \infty$	$D - G - E, H - G - E, F - H - G - E$

**Financial Contagion** The two vulnerability measures,  $\bar{v}_i$  and  $\tilde{v}_i$ , allow us to characterize contagion, with  $\bar{v}_i$  measuring the vulnerability of the source of financial contagion, and  $\tilde{v}_i$  measuring the vulnerability of banks in the credit chain.

If we consider a negative shock to investment return to one bank, say  $C$ , the first thing we look at is  $\bar{v}_i$ , vulnerability to the investment risk, for each bank making an investment in equilibrium. If this shock hits bank  $i$  but  $\bar{v}_i \geq \frac{1}{8} \text{ }_3$ , bank  $i$  is still solvent, and therefore no default or contagion takes place. As a bank with low  $\bar{v}_i$  is less vulnerable to investment risk, default and contagion are less likely to happen if an investment shock hits this bank.

So, we only look at banks with  $\bar{v}_i > \frac{1}{8} \text{ }_3$ . These are the ones that potentially trigger default of its creditors. In this example, when  $\bar{v}_i < \frac{y + \frac{3}{8}}{2(y + \frac{1}{8})}$ , these are bank  $F$  and bank  $H$ .

Then, we look at how the default shock propagates through the network. Following Acemoglu, Ozdaglar, and Tahbaz-Salehi (2013)[1], we look at the situation where only one bank defaults because of investment risk. We use a simple assumption that a borrower of bank  $i$  pays back nothing when it defaults. In this case,  $\tilde{v}_i$ , the percentage total repayment to bank  $i$ , is simply  $1 - \frac{[\text{loan to bank } j]}{[\text{total loan of } i]} \times 100$ . Then, we can compare  $\tilde{v}_i$  with  $\bar{v}_i$  to determine whether there is contagion. Hence, bank  $j$  triggers default of bank  $i$  if (1) bank  $j$  is an important borrower of bank  $i$  and (2) bank  $i$  has a high  $\bar{v}_i$ . If default of a bank triggers default of its creditor(s), repeat this procedure to see if creditors of the defaulting banks will default. Repeat this procedure until we reach the end of the credit chain or there is no further default.

Following this algorithm, Proposition 10 solves the vulnerability of the system to financial contagion.

**Proposition 10** *In this eight-bank example, assume  $c_4 = 0$ . Financial contagion triggered by investment risk is summarized in Table 3.*

- (i) *The default of periphery bank C never triggers default;*
- (ii) *The default of bank D always triggers default;*
- (iii) *The default of bank G triggers default if  $y > \frac{1}{8} \text{ }_3$ ;*
- (iv) *The default of bank H triggers default if  $y > \frac{5}{8} \text{ }_3$ .*

When  $c$  is higher, the default of a borrower triggers a larger shock to the revenue of its creditors. Since

bank  $G$  and bank  $H$  are in the core, Proposition 10 shows the vulnerability of banks in the core to financial contagion. The core market maker,  $H$ , is not as vulnerable to financial contagion but is vulnerable to investment risk, because its own profit from investment is low. Bank  $G$ , on the other hand, who lends to two banks in the economy, is much more vulnerable to counterparty default.

## 6.2 Strategic Default and Endogenous Trading Capacity

So far, we have taken trading capacity  $N$  as given. In this section, we explore how such capacity is bounded by banks' incentive to default strategically, when they have limited commitment.

### 6.2.1 Implications on the Trading Capacity of the Unsecured Lending Market

To understand strategic default in the unsecured lending market, where repayment depends on banks' reputation, we extend our model to an infinite-horizon setup, so that the value of reputation is endogenous. Time is discrete and lasts forever. Each period represents a trading day in the interbank market. The parameter  $N$  therefore represents the number of trading rounds within a trading day.

The repayment of the lending contract has to be incentive compatible for the borrower. Banks borrow on their reputation. In the following subsections, we first will study the implications of the model on such variables as leverage, assuming that the incentive compatibility constraint to repay a loan is not binding. And the number of trading rounds in a day,  $N$ , is also set exogenously. Then we will look at how the reputation and collateral requirement may constrain interbank lending and affect market liquidity and the number of trading rounds within a day.

The period payoff at period  $t$  is  $W_{0t}(\cdot)$ , which is solved in the previous section, taking as given the number of trading rounds in period  $t$ ,  $N_t$ . So the value from participating in the interbank trading is,  $V_t(\cdot) = \sum_{s=t}^{\infty} \beta^{s-t} W_{0s}(\cdot)$ , where  $\beta$  is the discount factor.

With unsecured lending, banks' incentive to repay depends on the value of reputation, which is other banks belief that the bank will not default. We assume that the reputation of a bank is public knowledge. If a bank defaults, we assume that its reputation is destroyed forever, and the bank will live in Autarky after the default. We focus on the stationary equilibrium.

A bank's continuation value in autarky is  $U(\cdot) = \frac{y^a}{1-\beta}$ . In the equilibrium, repayment is only incentive compatible if the payoff from default,  $D(\cdot) + \beta U(\cdot) + X$ , is no greater than the value from not defaulting,  $V(\cdot) + X$ .<sup>20</sup> So, incentive compatibility implies that,

$$D(\cdot) \leq [V(\cdot) - U(\cdot)].$$

---

<sup>20</sup> $X$  denotes the payoff from the bank's asset, which includes lending and investment.

$D(\cdot)$  is an increasing function in  $N$  and  $[V(\cdot) - U(\cdot)]$  is a decreasing function in  $N$ . So there exists a maximum number of trading rounds such that all banks still have incentive to repay their debt.

From Proposition 9, banks of low volatility type build up higher debt holding from market making activities and have less gain from participating in the game, the maximum depends on their incentive to default.

This upper bound for the maximum number of trading rounds, depends on endogenously on dealers incentive to default and gain from market making. When the market is less liquid, dealers could make more profit. This gives dealers more incentive to avoid default and maintain a good reputation. Meanwhile, competition among dealers reduces their profit margin and increases market liquidity. In equilibrium, a balance is reached between competition, market liquidity and dealers' incentive to maintain their reputation.

**Proposition 11** *There exists a maximum market liquidity parameter,  $N^*$ , such that all banks in the unsecured lending market have incentive to repay. The maximum is increasing in the dispersion of volatility type distribution.*

The result is driven by a trade-off between efficiency gain of liquidity to improve funding allocation and higher debt obligation built up through market making. When the volatility type distribution is more dispersed, the efficiency gain from improving allocation increases. So  $N^*$  increases.

## 6.2.2 Implications on the Trading Capacity of the Collateralize Lending Market

With collateralized lending, banks' incentive to repay depends on the value of collateral they pledge. Suppose the value of collateral each bank holds is  $Q$ . Then the incentive compatibility constraint implies

$$D(\cdot) \leq Q,$$

which imposes an upper bound on the trading capacity.

**Proposition 12** *There exists a maximum market liquidity parameter,  $N^*$ , such that all banks in the secured lending market have incentive to repay. The maximum is increasing in the dispersion of volatility type distribution, decreasing in the riskiness of collateral asset.*

## 7 Conclusion

In this paper, we build a dynamic model of an over-the-counter market, in which market making activities and a network structure emerge endogenously. The network structure is qualitatively similar to what we observe in a typical OTC market. The key mechanism behind these results is sorting on the volatility of traders' preferences over assets. Market-making services offered by traders with less volatile preferences

insure traders with more volatile preferences against their trading needs, which could be either selling or buying assets. The model gives us a fresh understanding of the economics behind the trading patterns in the OTC market. The policy implications 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] Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2013). Systemic risk and stability in financial networks. Technical report, National Bureau of Economic Research.
- [2] Afonso, G., A. Kovner, and A. Schoar (2013). Trading partners in the interbank lending market. *FRB of New York Sta Report* (620).
- [3] Afonso, G. and R. Lagos (2012). An empirical study of trade dynamics in the fed funds market. *FRB of New York sta report* (550).
- [4] Afonso, G. and R. Lagos (2014). Trade dynamics in the market for federal funds. Technical report, National Bureau of Economic Research.
- [5] Allen, F. and D. Gale (2000). Financial contagion. *Journal of political economy* 108(1), 1–33.
- [6] Atkeson, A. G., A. L. Eisfeldt, and P.-O. Weill (2014). Entry and exit in otc derivatives markets. Technical report, National Bureau of Economic Research.
- [7] Babus, A. (2007). The formation of financial networks.
- [8] Babus, A. (2012). Endogenous intermediation in over-the-counter markets. *Available at SSRN 1985369*.
- [9] Babus, A. and P. Kondor (2013). Trading and information diffusion in over-the-counter markets.
- [10] Bech, M. L. and E. Atalay (2010). The topology of the federal funds market. *Physica A: Statistical Mechanics and its Applications* 389(22), 5223–5246.
- [11] Chiu, J. and C. Monnet (2014). Relationship lending in a tiered interbank market. working paper.
- [12] Duffie, D., N. Gârleanu, and L. H. Pedersen (2005). Over-the-counter markets. *Econometrica* 73(6), 1815–1847.
- [13] Farboodi, M. (2014). Intermediation and voluntary exposure to counterparty risk. *Available at SSRN 2535900*.
- [14] Gale, D. M. and S. Kariv (2007). Financial networks. *The American Economic Review*, 99–103.

- [15] Gofman, M. (2011). A network-based analysis of over-the-counter markets. In *AFA 2012 Chicago Meetings Paper*.
- [16] Gofman, M. (2014). Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *Available at SSRN 2194357*.
- [17] Hojman, D. A. and A. Szeidl (2008). Core and periphery in networks. *Journal of Economic Theory* 139(1), 295–309.
- [18] Hollifield, B., A. Neklyudov, and C. S. Spatt (2012). Bid-ask spreads and the pricing of securitizations: 144a vs. registered securitizations.
- [19] Hugonnier, J., B. Lester, and P.-O. Weill (2014). Heterogeneity in decentralized asset markets. Technical report, National Bureau of Economic Research.
- [20] Jackson, M. O. (2005). A survey of network formation models: stability and efficiency. *Group Formation in Economics: Networks, Clubs, and Coalitions*, 11–49.
- [21] Kiyotaki, N. and J. Moore (2004). Credit chains. Technical report.
- [22] Lagos, R. and G. Rocheteau (2009). Liquidity in asset markets with search frictions. *Econometrica* 77(2), 403–426.
- [23] Lester, B., G. Rocheteau, and P.-O. Weill (2014). Competing for order flow in otc markets. Technical report, National Bureau of Economic Research.
- [24] Li, D. and N. Schürhoff (2014). Dealer networks.
- [25] Malamud, S. and M. Rostek (2014). Decentralized exchange.
- [26] Neklyudov, A. V. (2013). Bid-ask spreads and the decentralized interdealer markets: Core and peripheral dealers. Technical report, Working Paper, University of Lausanne.
- [27] Peltonen, T. A., M. Scheicher, and G. Vuillemey (2014). The network structure of the cds market and its determinants. *Journal of Financial Stability* 13, 118–133.
- [28] Rubinstein, A. and A. Wolinsky (1987). Middlemen. *The Quarterly Journal of Economics*, 581–594.
- [29] Wright, R. and Y.-Y. Wong (2014). Buyers, sellers, and middlemen: Variations on search-theoretic themes. *International Economic Review* 55(2), 375–397.

# A Appendix

## A.1 Omitted Proofs

### A.1.1 Proof for Proposition 2

**Proof.** Define  $W(\cdot, \cdot) \equiv \Omega(\cdot, \cdot) - W(\cdot, \cdot)$ .

$$W(\cdot, \cdot) = \begin{cases} a[\cdot + (2p-1)\cdot] + W^0(\cdot) + W^0(\cdot) - W(\cdot, \cdot), & \text{for } \cdot > \cdot, \\ a[\cdot + (2p-1)\cdot] + W^0(\cdot) + W^0(\cdot) - W(\cdot, \cdot), & \text{for } \cdot \geq \cdot \end{cases}$$

By construction of  $W(\cdot, \cdot)$ , for any  $\cdot \in [\cdot, H]$ ,

$$W_2(\cdot, \cdot) = \begin{cases} 0, & \text{for } \cdot > \cdot, \\ [(2p-1) - 1]a = 2(p-1)a < 0, & \text{for } \cdot \geq \cdot \geq \cdot, \\ [(2p-1) - (2p-1)]a = 0, & \text{for } \cdot \geq \cdot > \cdot. \end{cases}$$

Moreover, for  $\cdot > \cdot > \cdot$  :  $W(\cdot, \cdot) - W(\cdot, \cdot) = W(\cdot, \cdot) - W(\cdot, \cdot) = 2a(1-p)(\cdot - \cdot) > 0$ .

Hence,  $\arg \max W(\cdot, \cdot) \in [\cdot_L, \cdot]$  for any  $\cdot \in [\cdot, H]$ .

Similarly, for any  $\cdot \in [0, \cdot]$ ,

$$W_2(\cdot, \cdot) = \begin{cases} 0, & \text{for } \cdot \geq \cdot, \\ 2(1-p)a, & \text{for } \cdot \geq \cdot > \cdot, \\ 0, & \text{for } \cdot \geq \cdot \geq \cdot. \end{cases}$$

And for  $\cdot > \cdot > \cdot$  :  $W(\cdot, \cdot) - W(\cdot, \cdot) = 2a(1-p)(\cdot - \cdot) > 0$ ,  $\arg \max W(\cdot, \cdot) \in [\cdot, H]$  for any

$\cdot \in [0, \cdot]$ . Lastly, one can see that this payoff satisfies the feasible within each pair (i.e.,  $W(\cdot, \cdot) + W(\cdot, \cdot) \leq \Omega(\cdot, \cdot)$ ):

$$\begin{aligned} W(\cdot_c) + W(\cdot_m) &= 2W(\cdot) + (1-2p)a(\cdot - \cdot_m) + (\cdot_c - \cdot)a \\ &= \Omega(\cdot_c, \cdot_m) \end{aligned}$$

■

### A.1.2 Proof for Proposition 3

**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  $\cdot_k$  to denote type-( $\cdot, k$ ). First, consider the case that  $f(\cdot_A, \cdot_A) > 0$  and  $f(\cdot_B, \cdot_B) > 0$



and Proposition 1 is not satisfied:  $A \geq A > \tilde{B} \geq \tilde{B}$

$$\begin{aligned}
& \Omega(A, A) + \Omega(\tilde{B}, \tilde{B}) \\
&= a [\max(A, A) - (1 - 2p_0) \min(A, A)] \\
&\quad + a [\max(\tilde{B}, \tilde{B}) - (1 - 2p_0) \min(\tilde{B}, \tilde{B})] \\
&\leq a (x_3 + x_4) - (1 - 2p_0) (x_1 + x_2) \\
&< a (x_3 + x_4) - (1 - 2p_1) (x_1 + x_2) = \Omega(A, \tilde{B}) + \Omega(\tilde{B}, A),
\end{aligned}$$

where  $x_i$  is the  $i$ th order statistic of  $\{A, A, \tilde{B}, \tilde{B}\}$ . Second, suppose that Proposition 1 is satisfied –that is,  $f(x_k, x_k) > 0$  only if  $x_k \in [0, x]$  and  $x_k \in [x, H]$  – but  $f(A, A) > 0$  and  $f(\tilde{B}, \tilde{B}) > 0$ ,

$$\begin{aligned}
& \Omega(A, A) + \Omega(\tilde{B}, \tilde{B}) \\
&= a (A + \tilde{B}) - (1 - 2p_0) (A + \tilde{B}) \\
&< a (A + \tilde{B}) - (1 - 2p_1) (A + \tilde{B}) \\
&< \Omega(A, \tilde{B}) + \Omega(A, \tilde{B}).
\end{aligned}$$

Lastly, consider the case that  $f(x_k, x_k) = 0$  (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 1 and  $f(x_k, x_k) = 0$  are satisfied. ■

### A.1.3 Proof for Proposition 4

**Proof.** As proved in Proposition 2, 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:

$$\begin{aligned}
& \Omega(x_k, \tilde{x}_k) - W(\tilde{x}_k) \\
&= a \max(x_k, \tilde{x}_k) - (1 - 2p^-) \min(x_k, \tilde{x}_k) - W(\tilde{x}_k) \\
&< a \max(x_k, \tilde{x}_{-k}) - (1 - 2p^+) \min(x_k, \tilde{x}_{-k}) - W(\tilde{x}_{-k}) \\
&= \Omega(x_k, \tilde{x}_{-k}) - W(\tilde{x}_{-k}) \leq W(x_k).
\end{aligned}$$

■

### A.1.4 Proof for Proposition 5

**Proof.** We now show that the constructed matching rules and participation decisions satisfy traders' optimality condition for any  $t$ . We first establish the following lemma to describe traders' participation decisions.

For any  $\sigma > \sigma_N$ , if a trader has reached his first best allocation for sure at period  $t$ , it is optimal for them to leave the market afterward. That is,  $W_t(\sigma) = W_t(\sigma_L, \sigma, 0) + W_t(\sigma_H, \sigma, 2a) = (y + \sigma)a$  for all  $t > t^*$ .<sup>21</sup>

**Lemma 2** For any  $\sigma > \sigma_N$ ,  $\{\emptyset\} = \arg \max_{\tilde{z}} \Omega_t(z, \tilde{z}) - W_t(\tilde{z})$  for all  $t \geq t^*$  if and only if  $W_t(\sigma) = W_t(\sigma_L, \sigma, 0) + W_t(\sigma_H, \sigma, 2a) = (y + \sigma)a$ .

**Proof.** First of all, observe that for any trader with volatility  $\sigma > \sigma_N$ , it is optimal for him to act as a customer at the last period, as  $W_N(\sigma) = (y + \sigma)a - T_N > [y + (2p - 1)\sigma]a + T_N$  for any  $\sigma > \sigma_N$ . Hence, given a trader has reached his first best at period  $t < N$ , the only potential gain from participating the market is to make the market for some periods and act as a customer at  $t \leq N$ . Given that he will reach his first best position at period  $t^*$  again, his payoff of such deviation is then given by

$$\tilde{W}(t) = (y + \sigma)a + \sum_{j=t}^{t^*} T_j - T \leq (y + \sigma)a + \sum_{k=t+1}^{t^*} T_k - T < (y + \sigma)a,$$

where  $j > t^*$  denotes the periods of market making that he chooses. Given  $\sum_{k=t+1}^{t^*} T_k - T = \Gamma(t) - \sum_{k=1}^{t^*} T_k < \Gamma(t) - \sum_{k=1}^{t^*} T_k - T = \Gamma(t) - \Gamma(t^*) \leq 0$ , the last inequality comes from the fact that  $\Gamma(t)$  is decreasing  $t$ . Therefore, it is optimal for such a trader to stop participating the market once he has reached his first best allocation.

Now, we show that if  $\{\emptyset\} = \arg \max_{\tilde{z}} \Omega_t(z, \tilde{z}) - W_t(\tilde{z})$  for all  $t \geq t^*$ , it must be the case that the trader has reached his first best allocation. Suppose not, the expected value of the type- $\sigma$  trader who acts as a market maker up to  $t^*$  and stops participating from  $t^* + 1$  is given by

$$\begin{aligned} \tilde{W}_0(\sigma) &= [y + (2p - 1)\sigma]a + \sum_{k=1}^{t^*} T_k < [y + (2p - 1)\sigma]a + \sum_{k=1}^N T_k \\ &\leq v(N, \sigma) - \Gamma(N) \leq W_0(\sigma) \end{aligned}$$

■

With lemma 1, we can show that

**Lemma 3** For any two traders  $z = (\sigma_2, 2a, A)$  and  $z = (\sigma_1, 0, B)$

$$\Omega_t(z, z) = (y + \sigma_2)a + \frac{1}{2} [W_{t+1}(\sigma_1, 0) + W_{t+1}(\sigma_1, 2a)],$$

where  $\sigma_2 = \max(\sigma, \sigma^*)$  and  $\sigma_1 = \min(\sigma, \sigma^*)$ .

<sup>21</sup>The notation  $W_t(\sigma, a)$  deviates from the convention that only observable variables are state variables. The abuse of notation is meant to differentiate a trader of observable type  $(\sigma, a)$ , whose preference type conditional the observables is either  $\sigma = c +$  or  $\sigma = c -$  with probability one, from a trader of observable type  $(\sigma, a)$ , whose preference type conditional on the observables is  $\tilde{\sigma}$ . The former is the case after a trader acts as a customer, trading with a market maker to receive efficient asset allocation.

In other word, the above lemma shows that, the maximized surplus of any two traders is reached when having the more volatile types reached his first best allocation and letting the less volatile type take on the misallocation and keep participating the market.

We now show that any trader  $\theta > \theta_t$  who remain active in period  $t$  will not have a profitable deviation by violating the matching rules. In the constructed equilibrium, those traders should be “customers” at round  $t$  and leave the market afterward. What is left to show is that two customers ( $\theta \geq \theta_t$ ) this period will not be matched with each other.

$$\tilde{W}_t(\theta) = (y + \theta_2)a + \frac{1}{2}[W_{t+1}(\theta_1, 0) + W_{t+1}(\theta_1, 2a)] - W_t(\theta) \leq (y + \theta_2)a - T_t \leq W_t(\theta)$$

Note that, the first inequality comes from the fact that for any  $\theta > \theta_t$ ,  $W_t(\theta) = y + \theta - T_t > W_{t+1}(\theta) + T_t$  for  $\forall \theta > \theta_t$ , and the equality holds only when  $\theta = 1$ . Hence, traders are strictly worse off when  $\theta < 1$  and there is no gain for such deviation when  $\theta = 1$ .

Second, consider traders  $\theta_N < \theta < \theta_t$ . Clearly, if a trader of type  $\theta$  is matched with a customer this period, he is strictly better off if he keeps trading, according to Lemma 2. So, what is left to show is that there is no profitable deviation if they choose to be matched with another trader of type  $\theta \in [\theta_t, \theta]$  this period.

$$\begin{aligned} \tilde{W}_t(\theta) &= (y + \theta_2)a + \frac{1}{2}[W_{t+1}(\theta_1, 0) + W_{t+1}(\theta_1, 2a)] - W_t(\theta) \\ &\leq W_{t+1}(\theta_2) + 2T_t + W_{t+1}(\theta_1) - W_t(\theta) = W_t(\theta), \end{aligned}$$

where  $(y + \theta_2)a \leq W_{t+1}(\theta) + 2T_t$  for  $\forall \theta < \theta_t$ . So, there is no profitable deviation for any “dealers” to be matched with another dealer at period  $t$ .

The last thing we have to check is that the trading surplus generated by any two traders in each period is large enough so that they always participate. The participation constraint with  $\theta \leq 1$  at period  $N$  is then given by:

$$\begin{aligned} [(y + \theta_2) + (y + (2p - 1)\theta_1)] &> 2y + (2p - 1)(\theta_1 + \theta_2) \\ \theta_2 [1 - (2p - 1)] &> (1 - \theta_1)[2y + (2p - 1)\theta_1] \\ (\theta, p) &\equiv \frac{-(2 - \theta_1)(2p - 1)}{(1 - \theta_1)} > \frac{2y}{N} \end{aligned}$$

$(\theta, p) \propto 2(1 - p) > 0$ . Hence, there exists  $\theta^- = \frac{2y + 2(2p - 1)\theta_N}{2y + 2p}$ , such that  $(\theta, p) > \frac{2y}{N} \forall \theta \in (\theta^-, 1]$ . We now show that under this assumption, all traders will participate according to the constructed equilibrium for any  $t < N$  as well. That is, trading gains next period is larger than the autarky value of two traders. That is, the following inequality must hold for any period  $t > 1$ :

$$\Omega_t(z, z) \geq [(y + \theta_2) + (y + (2p - 1)\theta_1)] > 2y + (2p - 1)(\theta_1 + \theta_2)$$

where  $\alpha_1 \in [\underline{L}, \alpha_t]$  and  $\alpha_2 \in [\alpha_t, \alpha_{t+1}]$ . Note that as long as this inequality holds for the smallest  $\alpha_2$  and largest  $\alpha_1$ , it will hold for  $\alpha_2 \in [\alpha_t, \alpha_{t+1}]$  and  $\alpha_1 \in [\underline{L}, \alpha_t]$ . Hence, the condition reduced to

$$\alpha_t [1 - (2p - 1)] > (1 - \alpha_t) [2y + (2p - 1) \alpha_t].$$

This condition is then guaranteed by the assumption, as  $(\alpha_t, p) > \frac{2y}{N} > \frac{2y}{t}$ . The participation at the first period is also guaranteed automatically for any  $p \geq 1/2$ ,

$$[(y + \alpha_2) + (y + (2p - 1) \alpha_1)] > 2y + (2p - 1)(\alpha_1 + \alpha_2) > 2y.$$

Hence, with this assumption, all traders will participate according to the constructed equilibrium. ■

### A.1.5 Proof for Proposition 7

**Proof.**  $\alpha_t(\cdot) \equiv \min\{t : \alpha_t(\cdot) = 0\}$ ,  $\beta_t(\cdot) \equiv \min\{t : \beta_t(\cdot) = 1\}$ . Denote the optimal allocation rule and participation rule to be  $\alpha_t(\cdot)$  and  $\beta_t(\cdot)$ .

**Claim 3** If  $\alpha_t \geq \frac{y}{y+2(1-p)} \frac{1}{N}$ ,  $\alpha_t(\cdot) = 0$  implies  $\beta_t(\cdot) = 1 \forall t < N$ ,  $\alpha_t \in \cdot$ . In other words, traders with misallocated assets only leave at period  $N$ .

**Proof.** With  $N$  rounds of trade, there are at least  $1/2^N$  measure of agents with misallocated assets left at round  $N$ , if the planner chooses not to force agents with misallocated assets to leave before period  $N$ . Then, there must be some agents of type  $\alpha_N$  left with misallocated assets. Then gain from keeping an agent of type  $\alpha_N$  with misallocated assets until period  $N$  is  $\alpha_N(y + \alpha_N)a - \alpha_N[y - (1 - 2p) \alpha_N]a + \alpha_N[y - (1 - 2p) \alpha_N]a$ . The gain from having the agent leave the gain at period  $t$  is  $\alpha_t[y - (1 - 2p) \alpha_t]a$ . The former must be greater than the latter for any  $\alpha_t$  and  $t$  if  $\alpha_t \geq \frac{y}{y+2(1-p)} \frac{1}{N}$ . Therefore, the planner chooses not to force agents with misallocated assets to leave before period  $N$ . ■

**Claim 4** The optimal choice of  $\alpha_t$  satisfy the following property:  $\alpha_t(\alpha_1) \leq \alpha_t(\alpha_2)$ , for all  $\alpha_1, \alpha_2 \in \cdot$  and  $\alpha_1 \geq \alpha_2$ . In other words,  $\alpha_t(\cdot)$  is weakly decreasing in  $\cdot$ .

**Proof.** If not, we can show that there exists a feasible deviation that improves welfare. If  $\alpha_t(\alpha_1), \alpha_t(\alpha_2) < N$ , then according to Claim 3,

$$\begin{aligned} & \alpha_t^{(\alpha_1)}(c + \alpha_1)y + \alpha_t^{(\alpha_2)}(c + \alpha_2)y \\ & \leq \alpha_t^{(\alpha_2)}(c + \alpha_1)y + \alpha_t^{(\alpha_1)}(c + \alpha_2)y. \end{aligned}$$

Similarly, we can welfare improving deviation by swapping the moment of leaving the game if  $\alpha_t(\alpha_2) = N$ . ■

$\alpha_t(\cdot, \cdot) \equiv \alpha_t(1 - \alpha_t(\cdot)) [2(1 - p) + (c + (2p - 1) \alpha_t(\cdot))]$ .  $\alpha_t(\cdot, \cdot)$  is the expected flow payoff from allocation to a trader of type  $\alpha_t$  at period  $t$ , if the participation rule is  $\alpha_t(\cdot)$  and  $\beta_t(\cdot) = \cdot$ .

**Claim 5**  $t(\cdot, \cdot)$  satisfies the single crossing property. If  $t(\cdot, \cdot) - t(\cdot, \cdot) > 0$  for  $\cdot > \cdot$ , then,  $t(\cdot, \cdot) - t(\cdot, \cdot) > 0$ , for  $\cdot > \cdot$ .

**Proof.**

$$t(\cdot, \cdot) - t(\cdot, \cdot) = {}^t \{(1 - (\cdot)) [2(1-p) + (c + (2p-1))] - (1 - (\cdot)) [2(1-p) + (c + (2p-1))]\}$$

If  $t(\cdot, \cdot) - t(\cdot, \cdot) > 0$ ,  $1 - (\cdot) > 0$ . Then for  $\cdot > \cdot$ ,

$$\begin{aligned} t(\cdot, \cdot) - t(\cdot, \cdot) &= {}^t \{(1 - (\cdot)) [2(1-p) + (c + (2p-1))] - (1 - (\cdot)) [2(1-p) + (c + (2p-1))]\} \\ &\geq {}^t \{(1 - (\cdot)) [2(1-p) + (c + (2p-1))] - (1 - (\cdot)) [2(1-p) + (c + (2p-1))]\} \\ &> 0 \end{aligned}$$

The first inequality uses Claim 4, that  $(\cdot)$  is weakly decreasing in  $\cdot$ . ■

**Claim 6** If  $t(\cdot) = 1$ , then  $t_{t+s}(\cdot) = 1$  for any  $s \geq 0$ .

**Proof.** From Claim 5, the constrained optimal allocation rule at  $t(\cdot)$  satisfies the following property: if  $t(\cdot) = 1$  then  $t(\cdot) = 1$ , for all  $\cdot \geq \cdot$ . The single crossing property implies that the constrained optimal allocation function,  $t(\cdot)$ , can be characterized cutoff rules about volatility type  $\cdot$ .

Then, if  $t(\cdot) = 1$  and  $t_{t+s}(\cdot) = 0$ , for some  $s > 0$ , there must exist a  $\cdot < \cdot$ , such that  $t(\cdot) = 0$  and  $t_{t+s}(\cdot) = 1$  for  $0 \leq s \leq s$ . Or to say, traders of type  $\cdot$  trade with traders of type  $\cdot$  to help them reach the first best allocation. Since  $\cdot < \cdot$ , this choice weakly reduces welfare. ■

**Claim 7**  $(\cdot) = (\cdot)$ , a trader stops participating once he receives efficient asset allocation.

**Proof.** From Claim 3,  $(\cdot) \geq (\cdot)$ . From Claim 6, once a trader receives efficient asset allocation, his asset allocation stays efficient. This means that he will not match with another trader once he reaches efficient asset allocation. Therefore, welfare increases if the planner lets the trader of type  $\cdot$  stop participating and receive payoff from asset allocation at  $t = (\cdot)$ . ■

**Claim 8** Traders of type  $\cdot \in (\cdot, \cdot]$ , for all  $t \in \{1, \dots, N\}$ , receive first best allocation at period  $t$  and participate in the trading game until period  $t$ .

**Proof.** From Claim 5, the set of traders who received first best allocation in period  $t$  is characterized by a cutoff volatility type. Because of the feasibility constraints imposed by bilateral matching, at most one half of those traders with misallocated assets at period  $t$  can start receiving first best allocation at period  $t$ . This means that the cutoff type at  $t$  is  $\cdot_t$ , such that  $G(\cdot_t) = 1/2^t$ .

Then, according 7, traders of type  $\cdot \in (\cdot, \cdot]$ , who start receiving efficient asset allocation, participate in the trading game until period  $t$ . ■

Finally, according to arguments in the proof for Proposition 5, the planner cannot find profitable deviation even with a less restrictive assumption on the lower bound on  $\gamma$  than that in Claim 3. And that lower bound is  $\gamma^- = \frac{y + \frac{N(2\rho-1)}{N\rho}}{y + \frac{N(2\rho-1)}{N\rho}}$ . ■

### A.1.6 Proof for Proposition 9

**Proof.** In general, the debt holding of a bank is the amount of bidding contracts it signs when it is making the market plus the amount of asking contracts it signs when it is a customer bank. Although the expected trade volume of banks of the same volatility type is the same, their leverage varies. A market maker may fail to trade if they meet a customer who does not have the need to trade. Even if a market maker trades in every round, he switches between being a borrower and a lender in alternating rounds, as his asset holding swings between  $2a$  and  $0$ . The debt repayment also depends on the interest rate on the loan, which varies across trading rounds. To measure the leverage of the market, we first look at the maximum amount of debt banks take on.

Given the number of trading rounds,  $N_t$ , the maximum amount of borrowing a bank can take on is  $N_t a$ , if it keeps trading throughout  $N_t$  trading rounds. As banks with volatility type less than  $\gamma_{N-1,t}$  are active in the market in all trading rounds, banks with the largest amount of outstanding debt are among those banks. For a market maker to borrow, the amount of debt repayment is equal to the product of the bidding price and the amount of borrowed asset. For a customer bank to borrow, the amount of debt repayment is equal to the ask price times the amount of asset being borrowed. If  $N_t$  is an odd number, the maximum amount of debt of a bank is

$$D_{\max} = \max_{i=1}^{(N_t-3)/2} 2q_{2i}^b a + 2q_{N_t-1}^a a, q_1^b a + \max_{i=1}^{(N_t-3)/2} 2q_{2i+1}^b a + 2q_{N_t}^a a .$$

If  $N_t$  is an even number,

$$D_{\max} = \max_{i=1}^{(N_t-2)/2} 2q_{2i}^b a + 2q_{N_t}^a a, q_1^b a + \max_{i=1}^{(N_t-4)/2} 2q_{2i+1}^b a + 2q_{N_t-1}^a a .$$

■

### A.1.7 Proof for Proposition 10

**Proof.**

As  $\bar{\gamma}_F = 0$ , the default of  $C$  will not trigger default of  $F$ , or other banks. So in iteration 2, we do not need to consider the default chain starting from  $C$ . Since  $2y - \gamma_3 < 2y - \gamma_3/2$ , the default of  $D$  will trigger for sure the default of  $G$ . If  $0 < y - \gamma_3/8 \mp 2\gamma_1$ , or  $y > \gamma_3/8 \pm 2\gamma_1$ , the default of  $F$  will trigger the default of  $H$ . If  $y + \frac{1}{8}\gamma_3 < 2y - \gamma_3/2$ , or,  $y > \frac{5}{8}\gamma_3$ , the default of  $H$  will trigger the default of  $G$ . These results are summarized in Table 4.

Table 4: Contagion: Iteration 1

Bank ID	Expected Return	Chain of Defaulting Banks			
		$C$	$D$	$F$	$H$
		$\tilde{i}$			
A	$y - 4$	1	1	1	1
B	$y - 4$	1	1	1	1
C	$y + 4$	n.a.	1	1	1
D	$y + 4$	1	n.a.	1	1
E	$y - 3$	1	1	1	1
F	$y + 3$	0	1	n.a.	1
G	$y - 2$	1	$\frac{2y-3}{3y-7} \frac{3}{8}$	1	$\frac{y+\frac{1}{8} \cdot 3}{3y-\frac{7}{8} \cdot 3}$
H	$y \pm 1$	1	1	0	n.a.

Note: In iteration 1, banks triggering defaults are those who may default because of investment risks.

Table 5: Contagion: Iteration 2

Bank ID	Expected Return	Chain of Defaulting Banks		
		$D - G$	$F - H$	$H - G$
		$\tilde{i}$		
A	$y - 4$	1	1	1
B	$y - 4$	1	1	1
C	$y + 4$	1	1	1
D	$y + 4$	n.a.	1	1
E	$y - 3$	0	1	0
F	$y + 3$	1	n.a.	1
G	$y - 2$	n.a.	$\frac{y+\frac{1}{8} \cdot 3}{3y-\frac{7}{8} \cdot 3}$	n.a.
H	$y \pm 1$	1	n.a.	n.a.

Table 6: Contagion: Iteration 3

Bank ID	Expected Return	Chain of Defaulting Banks	
		$D - G - E$	$F - H - G$
A	$y - 4$	0	1
B	$y - 4$	1	1
C	$y + 4$	1	1
D	$y + 4$	n.a.	1
E	$y - 3$	n.a.	0
F	$y + 3$	1	n.a.
G	$y - 2$	n.a.	n.a.
H	$y \pm 1$	1	n.a.

If  $y > 3/8$ , the default chain  $D - G$  will trigger the default of  $E$ . Likewise, the default chain  $H - G$  will trigger the default of  $E$  if  $y > 3/8$ . If  $2y - 3/2 > y + \frac{1}{8} - 3$ , or  $y > \frac{5}{8} - 3$ , the default chain  $F - H$  will trigger the default of  $G$ . These results are summarized in Table 5.

Since  $\bar{\pi}_A = 0$ , the default of bank  $E$  will not lead to further default. The results on the last round of iteration is summarized in Table 6. ■