Financial Market Structure and Risk Concentration*

Briana Chang† Shengxing Zhang‡

First Draft: January 2021, This Draft: August, 2023

Abstract

We present a framework that jointly determines trading networks and risk allocation among banks. Banks are ex-ante homogeneous and risk-averse, but their marginal costs of bearing risks may be diminishing. The optimal trading network is determined by the tradeoff between risk sharing vs. concentration, which in turn determines the level of aggregate risk exposures, amount of intermediate trades, dispersion of transaction prices and distribution of risk exposure across financial institutions. We show that a continuous change in fundamentals or regulations may lead to a structural change in the market structure, causing discontinuous changes in these observables.

Keywords: Network, Over-the-Counter Market, Regulations

JEL classification: C70, G1, G20

*We would like to thank Manuel Amador, Scott Bauguess, Piero Gottardi, Peter Kondor, Cecilia Parlatore, Guillaume Rocheteau, Marzena Rostek, Gideon Saar, Neil Wallace, and Pierre-Olivier Weill for their useful discussions and comments. We also thank seminar participants at Wharton, Penn State, LSE, UT Austin, Houston, Toronto, the Wisconsin School of Business, the Carlson School of Management, the Federal Reserve Bank of Minneapolis, 2022 SITE - Dynamic Games, Contracts, and Market, Philadelphia Fed Conference – the Impact of Post-Crisis Regulation on Financial Markets, AFA 2022, SED 2022, SAET 2022, and the Virtual Finance Theory Seminar.

†University of Wisconsin–Madison, Wisconsin School of Business, Finance Department, 975 University Avenue, Madison, WI 53706, USA; briana.chang@wisc.edu.

‡London School of Economics, Peking University HSBC Business School; szhang@phbs.pku.edu.cn.

1 Introduction

In this paper, we present a framework that jointly determines interbank trading links and risk allocations through them. Even when bankers are ex ante homogeneous and risk averse, they could trade not only to share risks but also to concentrate them, creating an asymmetric interbank network consistent with empirical regularities: a few banks have large balance sheets, exhibit large gross trading volume, and bear more risks in their asset positions. Our tractable framework delivers insights into how and when the interbank network and risk exposure of large banks may change, which is useful for evaluating the responses of the market structure to regulations and for understanding different market structures for assets with varied riskiness.

In our model, banks are averse to uncertain asset positions, and their initial asset positions are subject to idiosyncratic shocks. As is standard, interbank trade serves to diversify and reallocate banks' risky asset positions. We assume that decentralized trades are subject to limited information: banks only observe other banks' asset positions after they decide to match. Such frictions thus prevent banks from perfectly and immediately sharing their risks, capturing the spirit of search frictions in OTC markets.

Specifically, before observing their realized initial asset positions, bankers are committed to with whom, when, and how they trade through multiple (finite) rounds of bilateral trades. We require that at any point in time, bilateral matchings and terms of trade with their trading counterparties in current and future trading rounds be stable, thus allowing multiple sequential pairwise deviations. The collection of banks' counterparties and trades over all trading rounds can thus be interpreted as the ex ante trading network that banks form to overcome the underlying frictions: limited information and a finite number of counterparties.

A bank's final payoff depends on the riskiness of its asset position after bilateral trades in the OTC market. We allow for a general payoff function and analyze how it affects the equilibrium network. When banks' private marginal cost of bearing risks is diminishing in the risk level, the standard risk-sharing strategies can be suboptimal even though banks are risk averse and ex ante identical. Instead, banks may concentrate the risk exposure on one side of bilateral matches. The diminishing marginal cost of bearing risks naturally arises in many applications, such as when banks have limited liability or have options to improve their risk-bearing capacity or trading technologies.

The underlying network is thus generally determined by the tradeoff between risk concentration and risk sharing with a diminishing marginal cost of bearing risks. When the benefit of risk concentration is strong, the network thus becomes asymmetric, featuring a few banks that hold a disproportional amount of risks, which we refer to as the core banks.

We establish that, when banks have a diminishing marginal cost of bearing risks, the interbank network concentrates risk via positive assortative matching (PAM). That is, riskier banks, which bear more risks through past transactions, are matched with riskier banks. Our sequential approach combined with PAM admits a simple and tractable characterization of network formation (i.e., the joint determination of links and asset allocations), which allows us to formally analyze how the network response to the underlying parameters (such as asset riskiness, taxes or subsidies).

Despite all banks being ex-ante homogeneous, the endogenous market structure in our model determines the level of the aggregate risk exposure, amount of intermediate trades, dispersion of transaction prices and banks' risk-bearing capacities. Importantly, the degree of heterogeneity in transaction prices and banks risk-bearing capacities comove together and can exhibit a discontinuous change when there is a shift in the trading structure. This is consistent with the recent finding in Eisfeldt, Herskovic, and Liu (2022) that the cross-sectional dispersion of dealers' risk bearing capacity comoves strongly with interdealer price dispersion, and the heterogeneity varies significantly over time.

We analyze the unique implications associated with structural changes. We first show that a small increase in the balance sheet cost of holding the asset, either because of tightening regulation or because the asset becomes riskier, can result in a regime shift in the interbank network, whereupon banks switch from fully symmetric and sharing risks with each other to concentrating risks in a minimum set of core banks. In other words, banks systematically change their trading behavior through the interbank network, which results in a few extremely risky banks and jump increases in aggregate risks. In this sense, a small shock can trigger systemic risks through the interbank trading network.[1]

[1]Endogenous default risks in this application differ from those in standard theories of financial contagion. Standard theories take the interbank network as given and analyze how network amplifies and dampens the propagation of default risks ex post. We study equilibrium network formation and highlight the ex ante aggregate default risks.

Next, we analyze the effect of varying core size on the market structure and apply it to an environment where banks have the option of investing in better risk management or faster trading technologies. We establish the property of delayed risk concentration, which uniquely pins down the time of connections given any size of core banks (i.e., those that invest in faster trading technologies). The optimal network is then reduced to choosing the optimal core size, trading off the cost of bearing risks versus the entry fee.

Core size as a summary statistic allows us to derive positive and normative implications of reforms that promote central clearing and/or discourage risk taking (such as post-2008 banking regulations), while accounting for the equilibrium response of the underlying market structure. Consistent with empirical evidence, our model predicts that policies that increase balance sheet costs relative to the entry fee could result in disintermediation and a more symmetric market structure. Moreover, we show that the effect of increasing banks' balance sheet costs on transaction costs becomes ambiguous whenever the underlying structure also changes, highlighting that transaction costs are generally not sufficient statics for assessing welfare.

Related Literature Methodologically, our dynamic framework with repeated bilateral matching[2] contributes a tractable approach to studying the formation of a trading network. It differs from existing approaches in the network formation literature[3] because it breaks down a complex network formation game into a sequence of subgames, each of which involves one round of bilateral matching together with asset trading, and a subsequent subgame. How an agent traded in the past is summarized by her characteristic, which becomes the state variable governing how she trades in later periods. By imposing sequential rationality, we can solve the network formation problem through backward induction.

[2]Most works in the matching literature involve a static environment, with only a few exceptions. Corbae, Temzelides, and Wright (2003) introduced directed matching into the money literature, where the key state variable is agents' money holding. Because there are no information frictions in Corbae, Temzelides, and Wright (2003), belief updating is not essential for their analysis, whereas it is a key component of our theory. With regard to the labor market, Anderson and Smith (2010) analyzed the dynamic matching pattern for which the public belief about an agent's skill (i.e., her reputation) evolves according to matching decisions. In our trading environment, the updating process depends endogenously on both agents' matching decisions and the terms of trade within a match.

[3]See the survey in Jackson 2005 for overview. Specifically, papers that have studied network formation in the financial market include Hojman and Szeidl (2008), Gale and Kariv (2007), Babus and Hu (2017), Cabrales, Gottardi, and Vega-Redondo (2017), Farboodi (2014), and Wang (2016)), where the last two papers in particular focus on the core-periphery structure.

While we use pairwise stability to characterize the equilibrium matching in a subgame, a deviating agent in a subgame can change all her future links, not just one link as in the static setup usually adopted in the literature . This method derives a unique solution. It is thus in sharp contrast to the standard network formation problem where agents simultaneously form multiple links, which is often subject to the curse of dimensionality and prone to multiple equilibria, because pairwise stability allows for the deviation of only one pair of agents even though agents form multiple links.

A similar approach is used in our previous work, Chang and Zhang (2018), where we consider a pure bilateral OTC market with risk-neutral agents and an indivisible asset. This paper allows for risk-averse agents and unrestricted asset holdings, which allows us to analyze risk concentration within the network.

Popular approaches to modeling OTC markets are based on random matching (e.g., Duffie, Gârleanu, and Pedersen 2005) or exogenous networks.[4] Relative to the literature that takes the network as given, our model provides a formal analysis of how the underlying structure of the OTC market might respond to policies.

One of our applications regards determination of the bilateral trading network and platform access. Thus, our paper also sheds new light on the literature on the costs and benefits of centralized vs. decentralized markets.[5] Instead of focusing on the tradeoff between these two markets, we allow for nonexclusive participation in both markets and emphasize how the participation decision in the centralized market interacts with the structure of the bilateral OTC market. The paper is related to recent works that study the coexistence of these two venues and market fragmentation, including Dugast, Üslü, and Weill (2019) and Babus and Parlatore (2017). Our framework is designed to analyze the network response, and the results can be generalized to environments where agents can access multiple types of platforms.

[4]For example, see Gofman (2011), Babus and Kondor (2018), and Malamud and Rostek (2014).

[5]Specifically, existing studies (e.g., Malamud and Rostek (2014), Glode and Opp (2019), and Yoon (2017)) consider other dimensions such as price impact and asymmetric information. They show that OTC markets can be beneficial for certain types of agents. In our model, a centralized platform is assumed to be a superior trading technology but requires a higher participation cost.

2 A Model of Trading Network Formation

2.1 Model Setup

The economy lasts $N+1$ periods, indexed by $t = 1, 2, \ldots, N+1$. It is populated by a continuum of banks of total measure 1, indexed by identity $i \in \mathbb{I} = [0, 1]$. Each bank is managed by a banker whose preferences and choices govern the bank. There are two types of consumption goods, numeraire goods and dividend goods, and one type of asset. The asset is a claim to a unit of dividend goods in each period.

A bank i receives a random initial asset position $a_{i,1}$, which is independently and identically distributed across banks, drawn from distribution $\pi_1(a)$. The randomness in the initial asset position can represent a liquidity shock that shifts the bank's asset position away from its ideal position. This could be the withdrawal of deposits by its customers. Alternatively, it could be a shock to a bank's valuation over an asset. We discuss the latter interpretation in greater detail after specifying the bank's preferences. Banks trade bilaterally from period 1 to period N. They have deep pockets in numeraire goods. In each period from $t = 1$ to $t = N$, there is a marketplace where banks meet bilaterally and exchange the asset with numeraire goods.

The preference of a bank i is

$$\mathbb{E}_1 \left\{ \sum_{t=1}^{N} \left[u_t(\widetilde{a}_{i,t}) - x_{i,t} \right] + u_{N+1}(a_{i,N+1}) \right\}, \tag{1}$$

where $\mathbb{E}_1$ denotes the expectation at the beginning of period 1, $u_t : \mathbb{R} \to \mathbb{R}$ is a concave utility function, $\widetilde{a}_{i,t}$ denotes the amount of dividend goods bank i consumes in period $t = 1, 2, \ldots, N$, which equals the bank's posttrade asset position in that period, and $x_{i,t}$ denotes the amount of numeraire goods the bank pays in exchange for assets.

The curvature in the utility function can be associated with the balance sheet costs of holding assets, which can be affected by regulations or the liquidity preference of a bank's depositors, for example. The heterogeneity in asset positions is the source of gains from trade because banks are risk averse. Transfers $x_{i,t}$ are a result of the transaction. $x_{i,t} = p_{i,t}(\widetilde{a}_{i,t} - a_{i,t})$, where $a_{i,t}$ denotes the pretrade asset position and $p_{i,t}$ denotes the price of the asset. The pretrade position in the following period, $a_{i,t+1}$, for $t = 1, 2, \ldots, N$, equals the posttrade asset position in period t, $\widetilde{a}_{i,t}$. Thus, the consumption of dividend goods in period $N + 1$ equals the posttrade asset position in the penultimate period, $a_{i,N+1} = \widetilde{a}_{i,N}$.

General preferences that depend on both preference shocks and asset endowment shocks, as in the frontier model in the literature on the OTC market, Üslü (2019), are allowed in our setup. Agents' flow utility from asset position $a_{i,t}$ in Üslü (2019) is $-\varepsilon_{i,t}a_{i,t} - a_{i,t}^2$, where $\varepsilon_{i,t}$ is an idiosyncratic preference shock. It is equivalent to $-(a_{i,t} - \bar{a}_{i,t})^2$ where $\bar{a}_{i,t} = -\varepsilon_{i,t}/2$. Thus, the preference shock, $\varepsilon_{i,t}$, is equivalent to a shock to the ideal asset position, $\bar{a}_{i,t}$. Agents receive shocks to their ideal asset positions rather than their asset holding as in our setting. If we regard the asset position in a bank's preference (1) as the deviation from the ideal position, the analysis in the rest of the paper applies to the more general setting.

Formation of Ex Ante Trading Network At the beginning of period 1, banks choose and commit to bilateral trading counterparties for periods $t = 1, 2, \ldots, N$ and their trading strategies with their counterparties. Denote the trading counterparty of bank i in period t $j_{i,t}$. The collection of bank i's counterparties $j_{i,t}$ over N rounds of trade forms its trading links. We assume that banks form their trading links unconditional on their realized asset holdings and valuations. Therefore, our setup effectively has a network formation stage ex ante. We can interpret trading links as permanent trading relationships between banks when we repeat the trading game with a fresh draw of shocks.

The assumption that banks form trading links ex ante that cannot be contingent on realized trading needs also avoids some technical complications in matching models under asymmetric information because trading needs can be banks' private information at the matching stage. Without this assumption and if trading needs were banks' private information, banks could in theory signal their types through different matching decisions, and the equilibrium would depend on how we specify off-equilibrium beliefs and require heavier notation. One could in theory impose off-equilibrium beliefs that support a pooling equilibrium and obtain the same outcome.

Contacting Frictions in Bilateral Trades Because trading counterparties are determined before banks receive shocks to their asset positions, bilateral trading counterparties are chosen subject to limited information, which prevents banks from locating ideal trading counterparties. Banks face uncertainty about their counterparty's asset position *before* contacting their counterparties in the corresponding period.

However, information is only limited at the matching stage in the beginning of period 1. Matched banks have complete information about each other's asset positions after they make contact, upon which they observe their counterparties' pretrade asset positions in the corresponding period.

If all banks could observe each other's realized positions before they chose their matches, the economy could achieve perfect risk sharing with one round of trade. For example, if banks' utility has a bliss point at 0, is symmetric around the bliss point, and the asset distribution is symmetric around 0, banks with position a would be matched with banks with the opposite position $-a$, and their posttrade positions would net out to zero (i.e., there would be perfect negative sorting on asset positions.) Hence, the assumed contacting frictions are intended to capture the spirit of conventional search frictions that prevent banks from locating their ideal trading counterparties.

Terms of Trade: Contingent Asset Flows and Prices While the connections are determined ex ante, trades depend on the realized asset positions of a bank and its counterparties, because trading takes place after it and its counterparties observe each other's realized asset positions within the match. Thus, if we regard the economy as a trading game within a trading day and repeat it over time, banks' realized asset positions change how they trade (i.e., the asset flows) within the network from day to day, although the network remains the same.

Formally, the terms of trade within a match, including both asset allocations and transfers of numeraire goods, are contingent on the realized positions of a bank i and its counterparty j, denoted by a_i and a_j, respectively. Let $y(i,j) = \{\widetilde{a}_k(a_i,a_j), \widetilde{x}_k(a_i,a_j), k \in \{i,j\}\}$ be the terms of trade within the match (i,j), where $\widetilde{a}_k(a_i,a_j)$ denotes the posttrade asset holding of bank k, and $\widetilde{x}_k(a_i,a_j)$ denotes the transfer to bank k, $k \in \{i,j\}$. The within-match transfers sum to zero:

$$\widetilde{x}_i(a_i,a_j) + \widetilde{x}_j(a_i,a_j) = 0. \tag{2}$$

The within-match asset allocation is feasible if

$$\widetilde{a}_i(a_i, a_j) + \widetilde{a}_j(a_i, a_j) = a_i + a_j. \tag{3}$$

The allocation of asset positions is associated with the allocation of risks from uncertain asset positions because given a distribution of banks i and j's pretrade asset positions, the posttrade positions also follow a distribution, which is the key characteristic that governs bilateral matching.

While the terms of trade are contingent on the realized positions within a pair, banks are committed to the contingent terms of trade ex ante. This is a strong assumption.

Sequential Choices of Trading Links and Terms of Trade Banks play a sequential game at the beginning of period one when they decide trading links and terms of trade ex ante: they make decisions for earlier trading rounds first. All trading links and terms of trade before a period t are public information when banks decide matching and within-match terms of trade for the period. They constitute the information set contingent on which banks' period-t strategies are chosen.

Note that the information set for period-t strategy does not include the realized trading history contingent on realized asset positions of a bank and its counterparties. This is consistent with our assumption that both the trading network and trading strategies are decided ex ante, not contingent on the realized trading history.

Bank i's strategy in period t includes the choice of her counterparty, $j_{i,t}$, and the terms of trade with the counterparty, $y_t(i, j)$ for $j = j_{i,t}$ conditional on the information set for that period. Period-t strategies are sequentially optimal given the common information set.

The common information set for period-t strategies can be summarized by the joint distribution of banks' asset positions. As we will see later, the gains from trade from period t onward depend on the trading history only through the joint distribution. We thus study a dynamic matching model with the joint distribution of banks' asset holdings and the marginal asset distribution as the evolving characteristics.

Evolving Characteristics Because banks' strategies are contingent on the public belief regarding the joint distribution of asset positions, characterizing its evolution over time is essential to our analysis. To understand how a bank's asset holding distribution evolves over time, consider the following example: suppose that a bank i bears all position exposures within its match in period 1. That is, its asset position in the next period equals the sum of it and its counterparty j's current asset positions, $a_{i,2} = a_{i,1} + a_{j,1}$. Denote the joint distribution of banks' asset holdings at the beginning of period t $\boldsymbol{\pi}_t : \mathbb{R}^{[0,1]} \to [0,1]$ and the marginal distribution of bank i's asset position at the beginning of period $\pi_{i,t}(a) : \mathbb{R} \to [0,1]$. Its posttrade asset distribution $\pi_{i,2}(a)$ now has mean zero and variance $2v_1$ when its pretrade position is uncorrelated with its counterparty's. On the other hand, under this first-period strategy, its counterparty's posttrade asset position is always zero, $a_{j,2} = 0$ (i.e., $\pi_{j,2}(a)$ is degenerate with both its mean and variance being zero).

The law of motion of the asset distribution of a bank i, $\pi_{i,t}(a)$, is given by Bayes' rule,

$$\pi_{i,t+1}(a) = \int\int \mathbb{I}(\widetilde{a}_{i,t}(a_i, a_j) \leq a)\boldsymbol{\pi}_{i,j,t}(da_i, da_j), \text{ for } a \in \mathbb{R}, \tag{4}$$

where $\boldsymbol{\pi}_{i,j,t}(a_i, a_{-i})$ denotes the joint distribution of bank i and its counterparty j's period-t pretrade asset positions. This again highlights the fact that bank i's posttrade asset distribution, $\pi_{i,t+1}(a)$, depends on the joint distribution of the pretrade asset positions of bank i and its optimally chosen counterparty and on how it trades with her counterparty, $\widetilde{a}_{i.t}(a_i, a_j)$.

2.2 Equilibrium Definition

Let $\Omega_t(i,j)$ denote the expected joint payoff between two matched banks in period t, i and j, given their equilibrium trading strategies. It maximizes their posttrade joint flow utility and continuation value given the pretrade aggregate distribution of asset positions in period t,

$$\Omega_t(i,j) \equiv \max_{\widetilde{a}_{i,t},\widetilde{a}_{j,t}} \mathbb{E}_1\left[u_t(\widetilde{a}_{i,t}) + u_t(\widetilde{a}_{j,t}); \left(y^t, j^t\right)\right] + \widehat{W}_{t+1}(i) + \widehat{W}_{t+1}(j) \tag{5}$$

subject to feasibility constraints, which depend on the pretrade joint asset distribution of banks i and j, $\pi_{i,j,t}(a_i, a_j)$. The within-match transfers do not appear in (5) because they sum to zero.

$\widehat{W}_{t+1}(i)$ denotes the bank's maximum payoff in the next period for the posttrade marginal distribution of bank i's asset holding $\pi_{i,t+1}(a)$ and the joint distribution with other banks' posttrade asset holding. Taking the aggregate distribution $\boldsymbol{\pi}_{t+1}$ and other banks' equilibrium payoffs $W_t(j)$ as given, the maximum payoff for a bank given a post-trade marginal distribution $\pi_{i,t+1}(a)$,

$$\widehat{W}_{t+1}(i) \equiv \max_j \Omega_{t+1}(i,j) - W_{t+1}(j) \tag{6}$$

for $t+1 \leq N$, where $W_{t+1}(j)$ is counterparty j's period-$t+1$ equilibrium payoff. For bank i that adopts equilibrium strategies until period t, its marginal asset position distribution equals its equilibrium marginal distribution, $\widehat{W}_{t+1}(i) = W_{t+1}(i)$.

The equilibrium in our model can be understood as competitive equilibrium in the literature on large games (McAfee 1993). Because there is a continuum of banks, a bank's decisions affect its own payoff, taking as given the aggregate distribution of matching and trading decisions in the market, and have a negligible effect on the aggregate distribution. A bank's deviating decision thus does not affect its counterparties' outside option and their payoff from the deviation. An agent's equilibrium payoff is the value of her outside option. This is analogous to the equilibrium market price in a competitive market that agents take as given. Thus, we alternatively call it her market utility.

Definition 1. Given the initial distribution of asset positions $\boldsymbol{\pi_1}$, an equilibrium consists of strategies $s^*_{i,t}$ for all $i \in [0,1]$, the market utility of agent i from period t onward $W_t(i)$ for all $i \in [0,1]$, and joint distribution of asset positions $\boldsymbol{\pi}^*_{t+1}$for all $1 \leq t \leq N$ such that the following properties hold:

1. Bilateral matches are stable. For any period $t \leq N$, if bank j is bank i's optimal counterparty, $j \in j_t(i)$, it solves (6), where the posttrade position $\{\tilde{a}_{i,t}, \tilde{a}_{j,t}\}$ maximizes Equation (5) and there is no profitable deviation when a measure ϵ of banks simultaneously deviate.

    (a) Feasibility of bilateral matching in any trading round $t \leq N$.

    (b) Dynamic Bayesian consistency: The joint asset distribution evolves following Bayes' rule given banks' strategies.

The equilibrium can be understood as multiple rounds of stable matching and trading. Our solution concept is stronger than the static pairwise stability solution concept. First, our sequential setting allows agents to deviate for multiple periods. When an agent and her counterparty initiate a pairwise deviation in a period t, they can switch their *subsequent* trading partners accordingly, provided that they promise their future counterparties at least their market utilities. Their counterparties can deviate. Second, we allow a simultaneous deviation among a sufficiently small measure ϵ of agents in period t. Therefore, the deviation in period t is not constrained by the equilibrium distribution of $\boldsymbol{\pi_t}$, so that deviating agents can flexibly create any joint distribution of asset positions among them. The minimal number of agents in the deviation in period t such that the deviation is flexible is 2^{N-t+1}. This allows agents in a joint deviation to form matches flexibly not only in the current period t but also in later periods. Since the measure of any finite number of banks is zero, the measure ϵ can be any positive value. We assume that the measure ϵ is sufficiently small that the effect of the group deviation on the aggregate distribution is negligible in our environment with a continuum of agents, and thus agents can take equilibrium market utility as given.

2.3 Equivalence and Uniqueness

We first show that the equilibrium outcome is unique and maximizes the aggregate payoff. Intuitively, this is because bankers have quasilinear preferences, have deep pockets in the numeraire goods and there is a continuum of traders. A competitive equilibrium in the large game is Pareto optimal. Under quasilinear preferences, any Pareto-optimal allocation maximizes the utilitarian social welfare.

Denote the aggregate payoff of the economy in period t by Π_t, which depends on the joint asset distribution $\boldsymbol{\pi}_t$. Given a strategy s_t in period t, the aggregate payoff equals

$$\Pi_t(\boldsymbol{\pi}_t) = \mathbb{E}_1 \int_0^1 u_t(\widetilde{a}_{i,t})di + \Pi_{t+1}(\boldsymbol{\pi}_{t+1}). \tag{7}$$

where $\mathbb{E}_1(u_t(\widetilde{a}_{i,t})) = \int\int u_t(\widetilde{a}_{i,t})\pi_{i,j_t(i),t}(da_i, da_{j_t(i)})$ and the terminal payoff is given by $\Pi_{N+1}(\boldsymbol{\pi}_{N+1}) = \mathbb{E}_1 \int_0^1 u_{N+1}(\widetilde{a}_{i,N+1})\,di$.

The following proposition shows that the equilibrium strategies – including agents' bilateral connections and the terms of trade within each match – maximize the aggregate payoffs.

Proposition 1. *Strategies* $\{s_{i,t}\}_{\forall i,t}$ *are equilibrium strategies if and only if they maximize* $\Pi_1(\boldsymbol{\pi}_1)$.

Proposition 1 has three implications. First, without any deviation between private and social values, the equilibrium is efficient.[6] Second, when a deviation arises for various reasons, one can implement the social planner's solution through taxes by simply aligning costs. Third, it implies that the equilibrium market structure and asset allocations through the market structure are payoff unique. The multiplicity that often makes it difficult to characterize financial networks does not arise in our framework. This provides the theoretical foundation for numerically solving the trading network.

Although the equilibrium is constrained efficient when taking traders' preferences as given, the equilibrium is socially optimal only if traders' private payoff is aligned with the social payoff. When there is a gap between the private payoff and the social payoff, we can use our framework to evaluate the divergence of the equilibrium market structure from the socially optimal structure.

3 Risk Distribution and Network Structure

Henceforth, we focus on more specific risk preferences for traders and analyze the resulting risk distribution and network structure.

Assumption 1. *For all trading rounds* $t \leq N$, *the flow utility is a quadratic function with bliss point at 0,* $u_t(a_{i,t}) = -\kappa_t a_{i,t}^2$, *where* $\kappa_t \geq 0$ *is a parameter for the flow cost in period* t.

Assumption 2. *(Risk Aversion) In the final period* $N+1$, *the expected payoff of bank* i, *denoted by* $W_{N+1}(v_{i,N+1})$, *is a decreasing function of posttrade risk exposure, where* $v_{i,N+1}$ *is the variance of* $\pi_{i,N+1}(a)$.

Assumption A1 can be understood as mean-variance utility from dividend goods, where the mean (i.e., ideal asset position of a bank) is normalized to zero.[7] Parameter κ_t

[6] Because agents have quasilinear preferences, this is equivalent to solving for Pareto optimal allocations.

[7] More generally, $u_t(a_{i,t}) = \kappa_{0,t} a_{i,t} - \kappa_{1,t} a_{i,t}^2$ for positive $\kappa_{0,t}$ and $\kappa_{1,t}$. Because $\kappa_{0,t}$ does not contribute to the heterogeneity in marginal utility, it is without loss of generality to set it to zero.

then represents the balance sheet cost of holding nonzero asset positions during trading period t, which can be associated with the riskiness of the asset.

Assumption A2 allows the expected terminal payoff to be a general decreasing function of the variance of the posttrade asset position. We assume a decreasing function to avoid trivial risk-taking behaviors. A1 and A2 together imply that holding risks is fundamentally costly for all banks, which captures the standard risk-sharing incentives.

Diminishing Marginal Costs of Bearing Risks Although all banks are risk averse, we show below that whether banks would engage in risk sharing crucially depends on the convexity of $W_{N+1}(v)$ – a convex $W_{N+1}(v)$ represents diminishing marginal costs of bearing risks. Note that the mean-variance utility for the terminal payoff can be nested as a special case in which the terminal payoff function $W_{N+1}(v)$ is linear in v thus represents a constant marginal cost of bearing risks. Below are two examples, which we discuss in greater detail in Sections 4 and 5, that naturally result in a convex terminal payoff function $W_{N+1}(v)$.

Example 1 (Limited Liability): When banks are protected by limited liability, the marginal cost of taking additional risks can be lower for riskier banks because they are more likely to default and thereby offload the cost of holding low asset positions to such external creditors as depositors. For example, if the utility from final consumption of dividend goods is a CARA utility function, $u(c) = 1 - e^{-c}$, the expected value given the variance of asset position v is

$$W_{N+1}(v) = \int [1 - \exp(\max(a_{N+1}, -D))] \, d\pi(a_{N+1}) \tag{8}$$

where $D > 0$ denotes the face value of debt that the bank owes to depositors. We normalize the expected value of a_{N+1} to 0 and assume that the expected position is fixed due to regulation. One can then verify that $W_{N+1}(v)$ can be decreasing and convex in v for $v \leq 2^N v_0$ for some value of debt face value D.

Example 2 (Optional Investment in Risk-Sharing Technologies): Banks in practice have options to improve their risk-bearing capacity by investing in superior but more costly trading technologies. Whenever there are economies of scale in such an investment decision, banks with more posttrade risk exposure are more likely to invest and thus in turn have lowered their marginal cost of bearing risks. For example, assume that banks' final period payoff is quadratic in their asset holding, $u_{N+1}(a) = -\kappa_{N+1}a^2$, and the banks can choose to invest in technology and reduce the marginal cost of bearing risks to $\eta\kappa_{N+1}$ for a fee ϕ,[8] their terminal payoff can be expressed as

$$W_{N+1}(v) = \max\{-\kappa_{N+1}v, -\phi - \eta\kappa_{N+1}v\}, \tag{9}$$

which is convex in v.

3.1 Risk Sharing vs. Risk Concentration in Bilateral Trade

In this section, we study the allocation given any match. Instead of working with asset allocations, we first simplify the analysis by showing that any optimal allocation of assets can be mapped to an allocation of posttrade variance within the pair and the relevant state variable can be reduced to the *variance* of the distribution $\pi_{i,t}(a)$, which represents a bank's risk exposure over time.

Variance Representation Within a match (i, j), the posttrade positions $\tilde{a}_k(a_i, a_j)$ depend on the realized positions of the two banks (a_i, a_j). Given any allocation rule, denote the variance of posttrade positions $\tilde{v}_k \equiv Var(\tilde{a}_k(a_i, a_j))$. The feasibility constraint of assets within the trade, Equation (3), can be rewritten as the following constraint that connects the pretrade and posttrade risks of the two agents within the pair:

$$\tilde{v}_i + \tilde{v}_j + 2\tilde{\rho}_{ij}\sqrt{\tilde{v}_i\tilde{v}_j} = V_{ij}, \tag{10}$$

where $V_{ij} \equiv Var(a_i + a_j)$ denotes the variance of the sum of pretrade positions and $\tilde{\rho}_{ij}$ denotes the correlation between the posttrade positions of two banks. Note that $\tilde{v}_k$ and $\tilde{\rho}_{ij}$ endogenously depend on the asset allocation that banks choose.

Lemma 1. *Under Assumptions 1 and 2, the optimal marginal distributions of the posttrade position for all banks have zero mean, and the posttrade positions for any two*

[8]For example, a bank can access a competitive centralized platform with probably $1 - \eta$, which allows her to reach the ideal asset position 0 by trading the asset at market price 0. The convexity of the terminal payoff function $W_{N+1}(v)$ holds more generally when a bank has multiple options for investment in faster trading technologies and the decision can be contingent on realized asset positions. We provide such an example in Section A.3 in the Appendix.

matched banks are perfectly positively correlated ($\tilde{\rho}_{ij} = 1$). *Moreover, the pretrade positions of any two matched banks in the efficient solution are uncorrelated.*

Under A1 and A2, banks' payoff decreases with the variance and mean of their asset positions; hence, it is optimal to keep the means of their posttrade positions at zero. Moreover, since increasing the correlation within the match reduces the total posttrade variance, it is thus optimal to set $\tilde{\rho}_{ij} = 1$.

Moreover, a positive correlation between the pretrade positions of two matched banks necessarily increases the variance of their total pretrade positions, which is the right-hand side of the feasibility constraint for variance allocation, Equation (10). Thus, all else being equal, it is optimal to match banks whose asset positions are not correlated (negative correlation is not available when banks trade optimally). In other words, it is not optimal to match two banks twice because the asset positions of any two previously matched banks are positively correlated.

Given that the asset positions for all agents are uncorrelated on the path, the sufficient static of an agent's characteristic is her pretrade variance $v_{i,t}$. In other words, $v_{i,t}$ is the state variable, and thus, we now use $W_t(v_{i,t})$ to denote the bank's maximum payoff given her characteristic $v_{i,t}$.

Simplified Problem: Allocation of Risks Within Matches The optimal asset/risk allocation within any match (i, j) can thus be reformulated as optimally choosing the share of the risks within a pair (i, j), given their pretrade variance V_{ij}, where bank i holds a share $\alpha_i \in [0, 1]$ of the total position, so that $\tilde{a}_i(a_i, a_j) = \alpha_i(a_i + a_j)$, and bank j holds a share $\alpha_j = 1 - \alpha_i$. With abuse of notation, we use $\Omega_t(V_{ij})$ to represent the joint payoff between any two agents with pretrade variance V_{ij} in period t. The optimal share $\alpha \in [0, 1]$ maximizes and thus solves the joint expected payoff

$$\Omega_t(V_{ij}) = \max_{\alpha \in [0,1]} \left\{ -\kappa_t \left(\alpha^2 + (1-\alpha)^2 \right) V_{ij} + W_{t+1}(\alpha^2 V_{ij}) + W_{t+1}((1-\alpha)^2 V_{ij}) \right\} \tag{11}$$

Given any match, the optimal risk allocation within the pair must satisfy the FOC from Equation (11). The share of total asset positions allocated to agent i can be expressed as

$$\alpha_i = \frac{\kappa_t - W'_{t+1}((1-\alpha_i)^2 V_{ij})}{\left[\kappa_t - W'_{t+1}((1-\alpha_i)^2 V_{ij})\right] + \left[\kappa_t - W'_{t+1}(\alpha_i{}^2 V_{ij})\right]}. \tag{12}$$

The optimal risk allocation in bilateral trade characterized in (12) illustrates the tradeoff between risk concentration and risk sharing. Given that agents are risk averse, they would like to lower the posttrade variance. Any concentration of risks ($\alpha \neq \frac{1}{2}$) is costly in the sense that it leads to higher posttrade pairwise variance, which is given by $(\alpha^2 + (1-\alpha)^2)\, V_{ij}$. Such a cost generally increases with banks' balance cost κ_t. In the special case when $W_{t+1}(v)$ is linear in v (or $\kappa_t \to \infty$), it represents no benefit (extremely high cost) of risk concentration; hence, the solution is standard risk sharing (i.e., $\alpha = \frac{1}{2}$).

On the other hand, when $W_{t+1}(v)$ is sufficiently convex (and low κ_t), there could exist an interior solution where $\alpha > \frac{1}{2}$. That is, it might be optimal to allocate more risks to one of the agents. The degree of concentration thus crucially depends on the property of $W_{t+1}(v)$, which endogenously depends on the optimal choice of counterparties in our framework for any $t < N$. In this sense, Equation (12) highlights the connection between the choice of asset allocation and agents' counterparties in our dynamic setup.

3.2 Risk Concentration Through Sorting

We now turn to study to the optimal choice of counterparties for an agent with risk position v_i at period t, which endogenously determines the value of $W_t(v)$ and can be expressed as

$$W_t(v_i) = \max_j \left\{\Omega_t(v_i + v_j) - W_t(v_j)\right\}, \forall t \leq N$$

where we use the fact that $V_{ij} = v_i + v_j$ because pretrade positions of any two matched banks are uncorrelated (Lemma 1). The proposition below first establishes the dynamic matching outcomes given any terminal payoff function $W_{N+1}(v)$.

Proposition 2. *(1) Full Risk Sharing and Random Matching: When the terminal payoff function $W_{N+1}(v)$ is concave in variance v, the unique trading network exhibits full risk sharing and the matching outcome is equivalent to random matching. (2) Positive Assortative Matching (PAM) on Risk Exposure: When the terminal payoff function $W_{N+1}(v)$ is convex in variance v, the optimal sorting outcome is PAM on variance v_t in any trading*

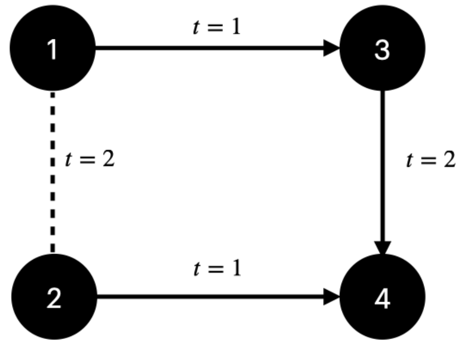

Figure 1: Risk-concentrating Network with PAM ($N = 2$)

round t. Within any match in which agents have the same variance v_t, the optimal share maximizes joint payoff $\Omega_t(2v_t)$.

With a concave terminal payoff function $W_{N+1}(v)$, there is no benefit of risk concentration. Hence, the solution is the standard risk sharing. Formally, one can show that $\alpha = \frac{1}{2}$ is the unique global maximum as the objective function in Equation (11) is concave in α. Given that agents share their exposure equally with any match over time, the evolution of variance thus yields $v_{i,t} = \frac{1}{2}v_{i,t-1} = \left(\frac{1}{2}\right)^{t-1} v_1 \forall i, t$. As there is no cross-sectional dispersion of $v_{i,t}$, the matching outcome is equivalent to random matching. In this sense, the trading outcome is the same as in Afonso and Lagos (2015), which can be nested in our framework as $W_{N+1}(v) = -\kappa_{N+1}v$.[9,10]

We focus on the case in which $W_{N+1}(v)$ is convex throughout the rest of the paper. With convex $W_{N+1}(v)$, one can show that, for any period t, $\Omega_t(V_{ij})$ is also convex in $V_{ij} = v_i + v_j$. Hence, given any distribution of $v_{i,t}$, agents are matched with counterparties that hold the same risk exposure; thus, on the equilibrium path, $V_{ij} = 2v_i$. Figure 1 illustrates an example of a network that features risk concentration with $N = 2$, where we use an arrow to point toward the agent with higher posttrade variance if asymmetric allocation arises within the match, and we use a dashed line to represent equal risk sharing. In this example. Agents 3 and 4 take on more risks from Agents 1 and 2 in period 1.

[9] Afonso and Lagos (2015) predicts that posttrade exposure is given by $a^k_{t+1} = \frac{a^i_t + a^j_t}{2}$, which implies that the posttrade variance is reduced to half, $v^i_{t+1} = \frac{v^i_t + v^j_t}{4}$. Since all agents share the risk equally, their characteristics remain the same ($v^i_t = \left(\frac{1}{2}\right)^t v_0 \ \forall i$).

[10] More generally, concavity in $W_{N+1}(v)$ predicts negative assortative matching (NAM). Even if the economy starts with two different initial values (say half of agents start with low (high) exposure v^L_0 (v^H_0)), all agents again become homogeneous in the subsequent periods under NAM.

In period 2, PAM implies that Agent 4 (Agent 1) is matched with Agent 3 (Agent 2). That is, an agent who obtains a higher share of risks from her counterparty in period $t-1$ then matches with another agent who also holds more risks from past transactions. Through this dynamic matching process, risks can be concentrated in a smaller set of agents within the network. Specifically, relative to random matching, positive sorting results in a higher level of concentration and a more convex $W_t(v)$.

Tractability of the Sequential Approach Our sequential approach combined ith PAM admits a simple and tractable characterization of network formation (i.e., the joint determination of links and asset allocations). The sequential choice of links allows us to break a network formation game into a sequence of bilateral matchings, where the effect of trading decisions before t is summarized by the characteristics of an agent (i.e., her risk position at period t, $v_{i,t}$). Bilateral matching (and trading) in period t determines their posttrade characteristics $v_{i,t+1}$, taking into account their future decisions. This effect is summarized by the continuation value.

The problem can thus be solved recursively via backward induction. Given $W_{t+1}(v)$ and PAM, the allocation within any pair v, denoted by $\alpha_t^*(v)$, is the solution to a one-dimensional optimization problem that maximizes $\Omega_t(2v)$ in Eq (11). The fact that matching is positive associative further simplifies the problem because the optimal solution is distribution-free, and the value function must satisfy $W_t(v) = \frac{1}{2}\Omega_t(2v)\ \forall t$. Given the policy functions $\alpha_t^*(v)$ and $W_t(v)$, one can then solve for the evolution of variances v_t^i over time.

Note that even when $W_{t+1}(v)$ is convex and thus the objective function might not be concave, the numerical algorithm remains computationally simple, as it still involves only one-dimensional optimization. Analytically, since there in general could exist multiple locally optimal solutions that satisfy the first-order necessary condition (12), we further impose conditions on $W_{N+1}(v)$ in Section 4 to provide full analytical characterization.

3.3 Price of Risks

While all agents are ex ante homogeneous, our model predicts dispersion of prices and risk-bearing capacities whenever the optimal network features asymmetric allocation (i.e., when the solution is not fully risk-sharing). Intuitively, since holding risks is costly, an agent within the pair that holds more posttrade variance needs to be compensated by receiving transfers from her counterparty. Through this channel, our model predicts heterogeneous prices across different trading pairs. In Figure 1, for example, Agent 1 will be trading with worse terms when trading with Agent 3 than with Agent 2. This holds even though there is no delay (i.e., $\kappa_t = 0$ for $t < N$ and $\kappa_N = 1$ and thus the environment is effectively static). We thus now formalize the expected transfer and implied transaction prices within each trading pair.

Expected Transaction Costs within Pair Given the optimal risk allocation within the match, the agent that holds lower risks post-trade must compensate the agent that holds more risks so that both agents are indifferent. Hence, the expected transfer within a pair of banks with variance v matched in period t, denoted by $x_t(v)$, solves

$$-\kappa_t \tilde{v}_t^l(v) + W_{t+1}(\tilde{v}_t^l(v)) - x_t(v) = -\kappa_t \tilde{v}_t^h(v) + W_{t+1}(\tilde{v}_t^h(v)) + x_t(v), \tag{13}$$

where the RHS (LHS) represents the payoff of the agent with lower (higher) posttrade variance $\tilde{v}_t^l(v)$ ($\tilde{v}_t^h(v)$).

Implementation with Bid-Ask Spreads. Note that the equilibrium transfer $x_t(v)$ within the pair can be implemented as a constant bid-ask spread times the expected trading volume. The bank that holds higher posttrade variance within the pair commits to bid and ask prices regardless of their realized asset positions, denoted $P_t^A(v)$ and $P_t^B(v)$, respectively. The bid-ask spread $S_t(v) \equiv P_t^A(v) - P_t^B(v)$ then solves

$$\left(\frac{S_t(v)}{2}\right) \vartheta_t(v) = x_t(v),$$

where $\vartheta_t(v) \equiv \mathbb{E}|\alpha_t^*(v)(a_{i,t} + a_{j,t}) - a_{i,t}|$ represents the expected volume between the pair of banks i and j.

Comovement in Dispersion of Prices and Risk-bearing Capacities Despite that all agents are homogeneous, the endogenous market structure determines the dispersion of prices and risk-bearing capacities in the market. As shown in Equation 13, a more asymmetric allocation within the pair v_t means that agents would have more heterogeneous risk-bearing capacities posttrade v_{t+1}. which then predicts a larger bid-ask spread within the pair v_t. Moreover, more heterogeneous v_{t+1} further predicts higher heterogeneous prices *across* pairs at $t+1$, as transfers generally depend on the characteristic of the pair.

Through this channel, our model naturally generates comovement in price dispersion in the interdealer market and the dispersion in dealers' risk bearing capacities, consistent with Eisfeldt, Herskovic, and Liu (2022). Moreover, Eisfeldt, Herskovic, and Liu (2022) also has shown that the heterogeneity varies significantly over time. Through the lens of our model, the degree of heterogeneity can change if the underlying risk-taking incentives triggers in different market structure, which we discussed in details In Section 4.

3.4 Evolution of Risks and Connections

Our sequential setting further offers predictions regarding how assets flow through the bilateral network and the evolution of risks across different banks. We now elaborate the mapping between the solution of our matching dynamic to the underlying network and establish how the risk evolution of an agent depends on her connections.

Mapping to Ex Ante Trading Network While all matching agents start with same pretrade variance, their risk positions could differ whenever $\alpha_t^*(v) \neq \frac{1}{2}$. Since there could be two different values of posttrade variance for each period and for any v_t, there are thus at most 2^N different paths of variance from period 1 to period N. The type of agents can thus be defined as a vector $\{\theta_{i,\tau}\}_{1\leq\tau\leq N}$, where $\theta_{i,\tau} \in \{h, l\}$ indicates whether agent i takes on higher ($\theta_{i,t} = h$) or lower variance ($\theta_{i,t} = l$) within her match in period τ. Since we assume one unit measure of agents, the measure of each type is thus $\frac{1}{2^N}$.[11]

The collection of an agent's trading links over N rounds as $\{j_t(i)\}_{\forall t\leq N}$ thus represents the ex ante network graph among 2^N types of agents, where $j_t(i)$ denotes the counterparty (direct link) for agent i in round t. The bilateral links among 2^N types of agents $j_t(i)$ $\forall i \in \{1, 2..., 2^N\}$ must satisfy PAM. For example, in Figure 1, since Agents 3 and 4 hold more risks in period 1 (i.e., $\theta_{3,1} = \theta_{4,1} = h$), PAM thus implies that $j_2(3) = 4$.

[11]Note that variance $\tilde{v}_t^h(v)$ is only weakly greater than variance $\tilde{v}_t^l(v_t)$. Hence, while there are 2^N types of agents, this definition also allows for agents to have the same realization. For example, in Figure 1, since Agents 1 and 2 share risk equally at period 2, these two agents will have the same path of v_t.

Time-varying Connections and Asset Flows One key difference of our sequential formulation from the standard network setting is that the set of agents that Agent i is "connected" with must be shrinking over time, since agents cannot use past links. For example, in Figure 1: Agent 1 is indirectly connected to Agent 4 in period 1, but she is no longer connected to Agent 3 nor Agent 4 at period 2.

Using the fact that the optimal allocations within the pair must take into account agents' future connections, Lemma 2 establishes a key property between these two. Under the optimal allocations, the risk-bearing cost of agent i in period t is one half of the harmonic mean of the posttrade risk-bearing cost of agent i and her counterparty.[12]

Lemma 2. *The marginal cost of holding risks for an agent with position v in period t is given by*

$$W_t'(v) = -\frac{1}{2}H\left(\kappa_t - W_{t+1}'\left(\tilde{v}_t^h(v_t)\right), \kappa_t - W_{t+1}'(\tilde{v}_t^l(v))\right) \ \forall t \leq N. \tag{14}$$

Consider again the example in Figure 1: Suppose that we refer to Agent 4 – who holds most risks at the end- - as the core agent. Equation (14) means that Agent 3 has a lower marginal cost of holding risk than Agent 1 in period 2, since she is connected to the core. Nevertheless, in period 1, Agents 1 and 3 must have the same risk-bearing capacity, as they will allocate the risks jointly accounting for both of their future connections.

This example also illustrates that agents' risk-bearing capacity will change over time. From the perspective of Agent 1, she has a lower marginal risk-bearing cost in period 1 since she is still indirectly connected to the core through Agent 3; however, after period 1, she loses this indirect access and thus will have a higher marginal risk-bearing cost in period 2.

Time-Varying Core Access As explained later in Section 4, whether an agent is connected to the core is often the sufficient static of an agent's risk-bearing capacity in period t. We thus now define the notion of connected agents in any period t and the concept of core access in our framework.

Formally, in any period t, an agent i is directly connected to counterparty $j_t(i)$, and thus is also indirectly to *future* counterparties of agent $j_t(i)$. Denote a set of agents I and their counterparties in period t $J_t(I)$, where $J_t(I) = \cup_{i \in I}\{i, j_t(i)\}$ is a list of period-t links

[12] The harmonic mean of any two variables γ_j and γ_j is $\frac{2}{\gamma_i^{-1}+\gamma_j^{-1}}$.

of agents in set I. The set of agents that agent i is connected to from period t onward can be understood as a tree with its root at the current match $J_t(i) = \{i, j_t(i)\}$.[13]

We thus let $\Psi_t(i) \equiv J_N(J_{N-1}(\ldots(J_{t+1}(J_t(i)))\ldots))$ denote the set of agents who are directly or indirectly connected to agent i in period t. By definition, an agent i is connected to at most 2^{N-t+1} agents from trading round t onward. In general, we regard the core agents as those whose final risk position $v_{i,N+1}$ is relatively high (i.e., above a certain percentile). Let $c_{i,N+1} = 1$ iff agent i is the core agent and $c_{i,N+1} = 0$ otherwise.

Definition 2. (Core Access) The core access of an agent in period t is given by

$$c_{i,t} \equiv \Sigma_{k\in\Psi_t(i)} c_{k,N+1}. \tag{15}$$

That is, in the example in Figure 1, where we refer to Agent 4 as the core, we thus have $c_{i,1} = 1$ and $c_{i,2} = 0$ for Agents 1 and 2. That is, these two agents no longer have core access in period 2. On the other hand, $c_{i,1} = 1$ and $c_{i,2} = 1$ for Agents 3 and 4.

Note that, by definition, $c_{i,t}$ must be (weakly) decreasing over time because the set of agents connected to agent i from period $t+1$ onward, $\Psi_{t+1}(i)$, is a subset of the set of agents connected to agent i from period t onward, $\Psi_t(i)$. As we will show later, agents with core access for a longer period will collect more risks over time, relative to the agents that lose core access earlier.

4 Structural Shifts

We now study how the underlying parameters affect the network and the implications of such changes. In Section 4.1, we start with an extreme case of structural shift, where banks switch from the standard full risk sharing (i.e., completely symmetric trades) to the maximum concentration with minimal core size. We use this result to show that such structural shift can trigger discontinuous increases in aggregate risks and prices despite having a rather smooth payoff function. In Section 4.2, we further consider a structural shift that changes core sizes.

[13] Due to the dynamic nature of our framework, future links are the specific factor that matters for current trading decisions. Thus, the relevant connections for an agent can be understood as a tree spanned from the current match. Nevertheless, the actual network does not need to be a tree. For example, according to Figure 1, the network graph contains loops.

4.1 From Risk Sharing to Risk Concentration

We now establish that a continuous change in the underlying parameters could trigger discontinuous change in aggregate risks and prices when agents move from full risk sharing to risk concentration. As discussed earlier, since the objective functions may not be concave, the full analytical characterization of the global optimum is not generally possible. We thus now show this result in two ways. First, to establish the comparative statics on the network analytically, we proceed by imposing additional assumptions on the payoff functions. Second, we provide an numerical illustration where the convexity of the payoff is driven by limited liability

Full Characterization We consider the environment where if risk concentration is optimal, it is without loss of generality to concentrate on at most *one* of the 2^N connected agents. In this sense, this is an extreme form of risk concentration. Moreover, to highlight the unique effect of interconnectedness, we further focus on the payoff function so that the corresponding policy function is continuous if there is only one round of bilateral trade. The sufficient conditions for these are formally stated in Assumption A3.

Assumption 3. *(1) The terminal payoff function* $W_{N+1}(v)$ is $k-$times continuously differentiable with bounded derivatives, has a marginal cost of bearing risk converging to 0 at infinite variance, $\lim_{v\to\infty} W'_{N+1}(v) \to 0$, and is strictly convex, $W''_{N+1}(v) > 0$; (2) function $\chi(v) \equiv \frac{1}{2}W'_{N+1}(v) + W''_{N+1}(v)v$ increases in v and is concave for all variance $v \in [0, 2^N v_1]$.

The first part of A3 guarantees that the solution is interior and convexity is bounded. Given that the benefit of concentration increases with the convexity $W''_{N+1}(v)$, and the cost of holding risks is captured by $W'_{N+1}(v)$, one can interpret the function $\chi(v) \equiv \frac{1}{2}W'_{N+1}(v)+W''_{N+1}(v)v$ as the relative benefit of risk concentration. Intuitively, the second part of A3 means that the benefit of concentration increases in v. As a result, it is optimal to have one core agent to continuously accumulate risks from others. Hence, having two core agents that have a moderate value of variance is dominated by having one core agent with a high value of variance. A concave $\chi(v)$ is a sufficient condition to guarantee that the objective function of Equation (11) for period N is single-peaked, and the corresponding solution $\alpha_N(v)$ is continuous.

Given A3 and $\kappa_t = 0$, we show that it is without loss of generality to concentrate on at most *one* of the 2^N connected agents. The allocation problem among 2^N agents can be greatly reduced to a one-dimensional problem, where the optimal risk allocation solves

$$\Pi(v_1) = \frac{1}{2^N} \max_{\alpha_s \geq \frac{1}{2^N}} \left\{ W_{N+1}\left(\alpha_s^2 \left(2^N v_1\right)\right) + \left(2^N - 1\right) W_{N+1}\left(\left(\frac{1-\alpha_s}{2^N - 1}\right)^2 \left(2^N v_1\right)\right)\right\}. \tag{16}$$

That is, the aggregate payoff can be understood as 2^N agents sharing a total risk of $V = 2^N v_1$ in a static environment. If all agents share risks equally, this means that $\alpha_s = \frac{1}{2^N}$, and thus the posttrade position for all agents yields $v^i_{N+1} = \left(\frac{1}{2^N}\right)^2 V = \frac{v_1}{2^N}\ \forall i$. That is, the posttrade variance for all agents decreases by a factor of $\frac{1}{2^N}$ after N rounds of risk sharing. On the other hand, $\alpha_s > \frac{1}{2^N}$ represents the case where there could exist one *core* agent who hold more risks than the remaining $2^N - 1$ agents.

Proposition 3. *Under Assumption A3 and $\kappa_t = 0$ for all $t \leq N$, there exists a cutoff v^* such that the equilibrium features full risk sharing when banks' initial risk exposure is below the cutoff, $v_1 \leq v^*$, and features risk concentration in one core bank among the connected banks when their initial risk exposure is above the cutoff, $v_1 \geq v^*$. When there are multiple rounds of trade ($N > 1$), the aggregate posttrade risk exposure, $\int v_{i,N+1} di$, and the prices of risk increases discontinuously at v^*.*

The proposition highlights that for small initial risk exposure $v_1 \leq v^*$, it is optimal for banks to use their network to share risks; consequently, we have less aggregate risk exposure. For higher initial risk exposure $v_1 \geq v^*$, it is privately optimal for banks to shift to a concentrated structure, where $\frac{1}{2^N}$ fraction of banks (i.e., the only core agent among 2^N interconnected banks) bears disproportionately large risks, resulting in greater aggregate risks. Importantly, only when $N > 1$ does the solution to Equation (16) exhibit discontinuous jumps at v^*.

Numerical Illustration: Application with Limited Liability We only impose Assumption A3 in order to establish the result analytically. We now relax A3 and apply our model to an environment where the risk-taking incentive results from limited liability. The Figure 2 illustrates an example of regime shift using the specification in Equation 8 with a normal distribution. Our result implies that banks might collectively use their

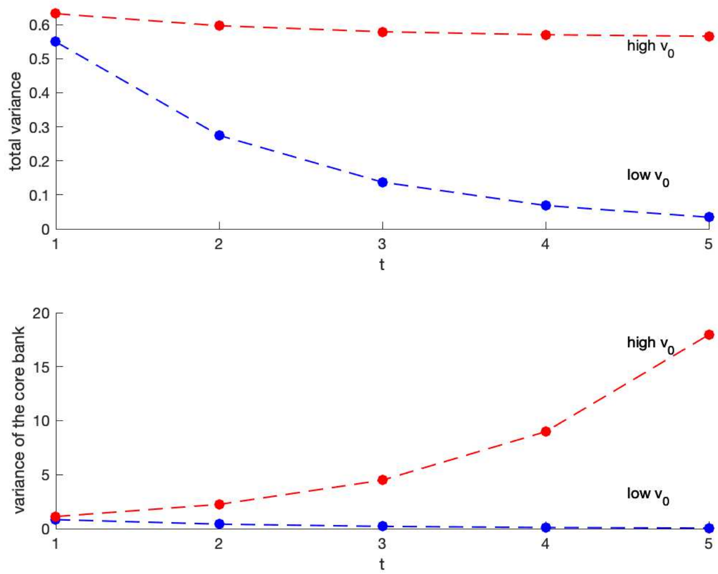

Figure 2: Regime Shift: $W_{N+1}(v) = -1 + e^{-cv}$, $c = 1.0196$, $v_{0L} = 1.02$ and $v_{0H} = 1.03$.

network to concentrate risks instead of sharing risks. Moreover, this result holds despite banks being risk averse (i.e., under Assumption 2).[14]

The red line represents the outcome where banks choose to share risks when $v_{0L} = 1.02$. Hence, each of them has low final risk exposure and default probability. The blue line, on the other hand, represents the case when banks collectively switch to concentrate risks on the core when $v_{0H} = 1.03$. As shown in the top figure, a small increase in the underlying uncertainty results in a jump increase in the aggregate probability of default (which is proportional to the total variance). In this sense, our model endogenously generates a "crisis" period. The bottom figure illustrates the risk positions of the core banks. As expected, since they start collecting risks from others during the crisis, their risk position increases over time instead of decreasing.

4.1.1 Flow of Risks Among Banks

We now elobarate how banks' risk positions invovle overtime when risk concentration arises (i.e., $\alpha_s > \frac{1}{2^N}$). Note that even though all non-core agents have the same terminal posttrade variance at the end, they could have have different evolution of risks positions

[14]Note that the standard risk-taking behavior arises when banks' payoffs are convex in their asset positions, and thus banks might prefer higher variance, which gives higher upsides. Our result here goes beyond this channel because we assume that $W_{N+1}(v)$ decreases in v.

over time, depending on their connections to the core. Hence, unlike a static model, our model gives predictions on how the asset flows within the network.

We now show that, given the optimal policy α_s that solves Equation 16, how each bank accumulates and/or unload risks from her counterparties over times can be characterized by the core access of an agent $c_{i,t}$ in period t. Formally, according to Lemma 2, when the flow utility in all trading rounds is zero, the risk capacity of an agent in period t is further reduced to $W_t'(v_{i,t}) = \left\{\Sigma_{k\in\Psi_t(i)}\left[W_{N+1}'(v_{k,N+1})\right]^{-1}\right\}^{-1}$. That is, the risk capacity of Agent i at period t is simply the harmonic mean of her connected counterparties' risk capacity in the terminal period $W_{N+1}'(v_{k,N+1})$. Specifically, $v_{k,N+1}$ could be two different values in this case, where $v_{N+1}^c = \alpha_s^2\left(2^N v_1\right)$ for the core agent, and $v_{N+1}^{nc} = \left(\frac{1-\alpha_s}{2^N-1}\right)^2\left(2^N v_1\right)$ for the non-core agents.

This thus shows that, if an Agent i is connected to the core in period t (i.e., $c_{i,t}=1$), then her risk capacity, denoted by $\gamma_t(c_{i,t}) \equiv -W_t(v_{i,t})$, is given by

$$\gamma_t(1) = \left(\left[-W_{N+1}'(v_{N+1}^c)\right]^{-1} + \left(2^{N-t+1}-1\right)\left[W_{N+1}'\left(v_{N+1}^{nc}\right)\right]^{-1}\right)^{-1}.$$

On the other hand, if an agent is not connected to a core, then

$$\gamma_t(0) = \frac{1}{2^{N-t+1}}\left[-W_{N+1}'\left(v_{N+1}^{nc}\right)\right].$$

Proposition 4. *Under Assumption A3 and $\kappa_t = 0$ for all $t \le N$, an agent with core access $c_{i,t} = 1$ within the pair collects $\frac{\gamma_{t+1}(0)}{\gamma_{t+1}(1)+\gamma_{t+1}(0)} > \frac{1}{2}$ share of the risks from her counterparties in period t. If neither of the agents within a pair has core access, they share risks equally.*

Recall that $c_{i,t}$ must be weakly decreasing over time, as $\Psi_{t+1}(i)$ is a subset of $\Psi_t(i)$. Each agent can be characterized by the period in which she loss core access, denoted by $\tau_i \equiv \sup\{N+1 \ge t \ge 1 : c_{i,t} = 1\}$. Intuitively, an agent holds more risks from her counterparty until she loses her core access, as formally established below.

Corollary 1. *An agent with core access until period τ_i collects risks for $\tau_i - 1$ periods, unloads her risks to her counterparty in period τ_i and then shares risks equally with her counterparties thereafter. An agent's expected trading volume increases with τ_i.*

In other words, if we use expected volume as the measurement of centrality, agents that have longer core access will become more central. In particular, an agent with direct connection to the core at earlier period in fact is less central (i.e., having lower expected volume) than an agent that has longer indirect core access. This highlights that the timing of connections matters more than the standard distance measure used in the network graph. Moreover, while these agents with longer core access have greater expected volume, they are not are not ultimately riskier because they eventually unload their risks to the core agent.[15] In this sense, having larger trading volume does not imply that they are riskier.

4.1.2 Predictions associated with the Rise of Risk-Concentration

A jump increase in transaction costs. If all banks adopt full risk-sharing, then our model predicts a zero bid-ask spread and thus the price across all pairs can be mapped to the fundamental price. When banks switch to risk-concentration, the bank within the pair must be compensated, predicting an increase in bid-ask spreads. Hence, as established in Proposition 3, it predicts a jump increase in transaction costs, even though the underlying asset characteristics remain similar.

A jump increase in dispersion of banks' risk-bearing capacities and prices. Structural shift from fully risk-sharing to risk-concentration also predicts a jump increase in interdealer price dispersion and risk-bearing capacities. Our model also provides an explanation for the finding in Eisfeldt, Herskovic, and Liu (2022). Through the lens of our model, the normal time (low price dispersion time) is the period where banks adopt risk sharing and thus relatively low price dispersion.[16]A slight increase in agents' risk-taking incentives can result in banks collectively switching to risk concentration, predicting a jump increase in interdealer price dispersion *leading* to the financial crisis.[17]

[15]In fact, as we discuss in the next section, when the flow cost of bearing risks is positive ($\kappa_t > 0$) for $t \leq N$, the noncore bank directly matched to the core in the final trading period in fact has the lowest posttrade risk exposure. In this sense, while these banks have been collecting risks over time and are "closest" to the core, they actually ultimately become the least risky.

[16]Note that our model predicts zero price dispersion when all banks adopt risk sharing, since we assume no other transaction costs and underlying frictions. Adding any frictions might then increase the baseline level of price dispersion. In this sense, our model focuses on explaining the spike (the difference) over time.

[17]Since our model is about the ex-ante risk exposures in the banking sector (instead of the ex-post contagion risks), we interpret this endogenous regime shift happens before the collapse.

4.2 Disintermediation with Increasing Core Size

We now analyze the possibility of varying core sizes and how such a change affects the distribution of risks and prices. We use this result to show that, in response to the regulation that increases the balance sheet cost of holding risks, the optimal network would now feature lower level intermediation along with an increase in the core size.

Specifically, we analyze an environment in which all banks can potentially take a costly binary action (such as invest or not) to reduce their risk-bearing cost in period $N+1$. This option naturally gives rise to a convex payoff function $W_{N+1}(v)$. Since the investing banks will hold more risks, we refer those banks as the core banks. That is, $c_{i,N+1} = 1$ iff a bank chooses to invest.

Assumption 4. *Piecewise linear with binary action. The terminal payoff function solves the optimal decision of accessing a faster trading technology:*

$$W_{N+1}(v) = \max_{c_{N+1}\in\{0,1\}} \left\{-\gamma_{N+1}(c_{N+1})v - \varphi_{N+1}(c_{N+1})\right\}, \tag{17}$$

where $\gamma_{N+1}(1) = \eta\gamma_{N+1}(0)$, $0 \leq \eta < 1$,and $\varphi_{N+1}(1) > \varphi_{N+1}(0)$.

Application: Platform Access One example of Assumption 4 is that banks can potentially pay a cost to obtain access to an exchange-like interdealer market.[18] Such a structure have been the focus of regulation and policy debates since the 2007-08 financial crisis, as many financial over-the-counter (OTC) markets show a classical two-tiered market structure where a few core banks have exclusive access to an exchange-like interdealer market.[19]

The example in Equation (9) can be nested as $\gamma_{N+1}(0) = -\kappa_{N+1}$, and $\gamma_{N+1}(1) = -\eta\kappa_{N+1}$, where the fixed cost of accessing the technology is denoted $\varphi_{N+1}(1) = \phi > \varphi_{N+1}(0) = 0$. The convexity of the terminal payoff function $W_{N+1}(v)$ holds more generally when a bank has multiple options for investment in faster trading technologies and the

[18] Note that while the timing of our framework implies that platform entry comes at the end, this assumption can be relaxed as long as there is a fixed cost associated with each entry. If there is no delay cost, it is indeed optimal to postpone access until the end, as agents would prefer to accumulate as much risk as possible from bilateral trades first before joining the platform.

[19] In particular, post-crisis reforms have increased dealer banks' balance sheet costs through tightened capital requirements and additional liquidity requirements and have promoted all-to-all exchanges. See detailed discussions in Yellen (2013) and Duffie (2018).

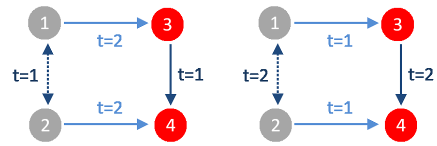

Figure 3: Late vs. early Concentration ($N = 2$)

decision can be contingent on realized asset positions. We provide such an example in Appendix A.3.

4.2.1 Dynamic Property of Core Access

When there are multiple cores, an additional question arises: Given the core size, what would be the optimal connections over time? To see this, consider Figure 3 where the total number of trading rounds $N = 2$ and thus four banks could potentially be connected. Suppose that the bilateral trading outcome under both networks is such that Banks 3 and 4 (Banks 1 and 2) have higher (lower) posttrade variance, given by $v_{i,N+1} = v_H$ for $i = 3, 4$ ($v_{i,N+1} = v_l$ for $i = 1, 2$). They, however, differ in terms of the timing of the bilateral connections. In the left graph of Figure 3, Bank 1 is first connected to Bank 2 and then Bank 3, but this order is reversed in the right graph.

Observe that the ordering of matching outcomes must result in different dynamic paths of $v_{i,t}$ despite having the same final outcome of $v_{i,N+1}$. Specifically, to concentrate risks on Agents 3 and 4, risk concentration takes place in period 2 for the left graph but in period 1 for the right graph. Since concentration necessarily results in higher total variance and is costly, it is optimal to delay risk concentration whenever the flow marginal cost of bearing risks in period $t \leq N$, κ_t is strictly positive. Thus, any solution that violates the backloading property is dominated.

Since risk concentration arises when agents have uneven core access, it is thus optimal to distribute core access as evenly as possible next period within a pair. This thus allows us to uniquely pin down the optimal dynamic connections to the core given any core size. For example, if Agents i and j are matched in period t and are connected to two core agents from that period onward, either of them maintains one core access from period $t + 1$ onward. That is, if $c_{i,t} = c_{j,t} = 2$, then $c_{i,t+1} = c_{j,t+1} = 1$,[20] which explains why the

[20]Recall that an agent's core access is defined in Equation 15, which must be (weakly) decreasing over time. The only difference is that $c_{i,t}$ can now be more than one, since an agent could be connected to

right graph cannot be optimal. This result can be established more generally beyond A4, which we relegate to the Appendix.

Lemma 3. *For any two agents i and j matched in period t, their posttrade core access is adjacent integers, $c_{i,t+1} = \lfloor \frac{c_{i,t}}{2} \rfloor$ and $c_{j,t+1} = \lceil \frac{c_{j,t}}{2} \rceil$. Under Assumption A4, the core access $c_{i,t}$ is the sufficient static for agent i's risk capacity in period t, where*

$$\gamma_t(c_{i,t}) = \frac{1}{2} H\left(\kappa_t + \gamma_{t+1}(\lfloor \frac{c_{i,t}}{2} \rfloor), \kappa_t + \gamma_{t+1}(\lceil \frac{c_{i,t}}{2} \rceil)\right) \ \forall t \leq N, \tag{18}$$

and $\gamma_t(c_{i,t})$ increase in $c_{i,t}$.

Under A4, the core access is again a sufficient static for agent $i's$ risk capacity. The only difference is that core access $c_{i,t}$ can be larger than one. As before, given $\gamma_t(c_{i,t})$, one can then pin down variance $\{v_{i,t}\}$ for all agents: Since $\gamma_t(c_{i,t})$ increases in $c_{i,t}$, agents who have more core access posttrade bear more risks.

4.2.2 Optimal Core Size

We have established that there is a unique optimal market structure given any core size c, which equals the initial core access for all agents connected from period 1 onward, $c_{i,1}$. The optimal network can then be further reduced to choosing the number of core agents in the beginning of the trading game among 2^N connected agents from period 1 onward. The expected ex ante payoff of an agent solves

$$\Pi(v_1) = \max_c \left\{ -\gamma_1(c) v_1 - \left(\frac{c}{2^N}\right) \phi \right\}. \tag{19}$$

Given any core size c, $\gamma_1(c)$ represents the risk-bearing capacity for all agents, taking into account future connections according to Equation 18. If there are c core agents among 2^N agents, the total measure of core agents would be $\frac{c}{2^N}$; hence, the second term captures the total entry costs.[21]

The tradeoff of core size can be seen from Equation 19: A larger core size results in higher total entry costs but lower risk-bearing costs. To explore how the core size

more than one core agent.

[21]Recall that an agent i can connect, directly or indirectly, to at most 2^N agents in N rounds of trade, where each type has a measure of $1/2^N$. Then, there are $1/2^N$ identical replicas of the finite network of size 2^N.

depends on the underlying parameters, we further assume the parameter for the flow cost of bearing risks, $\kappa_t = \delta\kappa_{N+1}\ \forall t \leq N$, where the parameter δ represents the flow cost of bearing risks in a trading period relative to the terminal period.

Proposition 5. *Under Assumption A4 and $\kappa_t = \delta\kappa_{N+1}\ \forall t \leq N$, given any (δ, η), the optimal core size is a weakly decreasing function of $\frac{\phi}{\kappa_{N+1}v_1}$.*

We prove this result by showing that, the risk capacity $\gamma_t(c)$ is a homogeneous function of degree 1 in κ_{N+1}. Recall that κ_{N+1} represents the balance cost of holding the assets and can be mapped to the riskiness of the underlying assets and v_1 represents the ex ante exposure. Hence, the ratio $\frac{\phi}{\kappa_{N+1}v_1}$ captures the entry cost *relative* to the level of risks. A higher value of the ratio means relatively higher costs of using the platform and thus a lower optimal core size.

Network Response to Regulatory Changes We now use Proposition (5) to shed light on the effect of polices that promote central clearing and/or discourage risk taking, following the adoption of post-2008 banking regulations. This includes policy that provide a subsidy for platform participation and/or increase taxes on banks' net exposure. Through the lens of model, the policy can be understood as increasing κ_t (i.e., making it more costly for banks to hold risks) and/or decreasing the entry cost of the platform (ϕ), both of which result in a lower $\frac{\phi}{\kappa_{N+1}v_1}$.

As predicted by Proposition 5, these polices induce an increase in participation in the central platform (i.e., a larger core size). As a result, there is also less risk concentration within the network, illustrated by Figure 4.

Our model predicts that the structure becomes more symmetric; nevertheless, the two-tier market structure persists. This explains why, as discussed in Collin-Dufresne, Junge, and Trolle (2018) and Duffie (2018), all-to-all trading has not materialized and the provision of clearing services remains concentrated.

4.2.3 Predictions associated with disintermediation

Lower dispersion of banks' risk-bearing capacities and aggregate volumes: As the size of cores increases, banks shift from risk-concentrating, market-making trades toward risk-sharing trades. Hence, the model predicts that banks' risk positions become less dispersed.

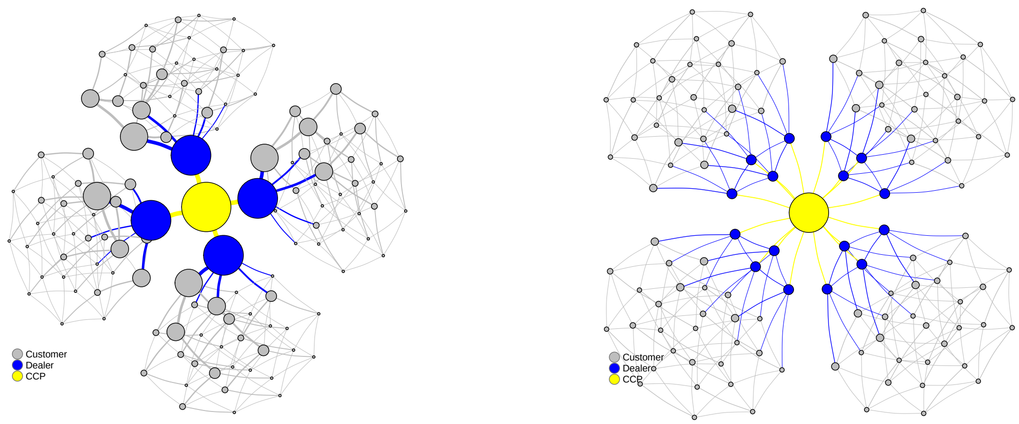

Figure 4: Pre vs. Post-regulation Market Structure.
Each panel shows the graph of the equilibrium trading network. In the network graph, each node represents a bank. The area of the node represents the gross trading volume involving the bank. The edges between nodes represent bilateral trading relationships. The width of an edge represents the bilateral trading volume. The left panel illustrates the pre-regulation market structure. The right panel illustrates the post-regulation market structure with increased balance sheet costs and lowered cost of accessing the centralized trading platform.

This is consistent with Eisfeldt, Herskovic, and Liu (2022), who documents Volcker rule (March 2014) resulted in lower dispersion of dealers' risk-bearing capacities.

Higher compositions of customer-to-customer trades: At a micro level, our model further predicts that "customers" trade more with other customers to hedge their risk exposures and less with big "dealers". To see this, as shown in the left panel of Figure 3, Agents 1 and 2, which we refer to as customers, now first trade with each other before meeting with the core dealers. In contrast to the case when there is only one dealer, they would unload risks to Agents 3 and 4 before meeting with each other. In other words, such change in the market structure means different timing of connections: customers now look for risk-sharing trades directly with other customers longer before unloading risks to dealers. Through this channel, our model predicts a higher volume between customer-to-customer connections, consistent with the empirical findings in Choi, Huh, and Seunghun Shin (2023). They show that customers increasingly provide liquidity following the adoption of post-2008 banking regulations.

Ambiguous effects on transaction costs: Since such a change in the market structure predicts higher volume of risk-sharing trades, it also has unique effect on the average transaction costs. Specifically, by looking at the weighted transaction costs, the average spread is thus given by $\frac{\Sigma_{t,v} S_t(v)\vartheta_t(v)}{\Sigma_{t,v}\vartheta_t(v)}$, where $\vartheta_t(v)$ represents the expected volume for any pair v at round t, one can see that such a structural change could result in lower average transaction costs despite the increase in the spread that market-makers charge. This is because that the spread between symmetric risk-sharing trades is zero and such a structural change results in higher shares of risk-sharing trades (i.e., C2C trades). This prediction is again consistent with Choi, Huh, and Seunghun Shin (2023), which shows the conventional bid-ask spread measures underestimate the cost of dealer's liquidity provision when C2C trades increase.

Indeed, our result highlights that the standard predictions – an increase in banks' balance sheet cost increases the bid-ask spreads and transaction costs – may not hold when the market structure also changes. This thus rationalizes the seemingly contradicting evidence in the post-Volcker rule era[22] and also implies that transaction costs are generally no longer a sufficient measure of welfare under an endogenous market structure.

Remark While both Section 4.1 and 4.2 can be driven by an increase in risks (measured by v_0) associated with an increase in the core size, they have different predictions. This is because that in Section 4.1, we focus on structural shift starting from fully risk-sharing, hence, an increase in v_0 predicts the rise of the core (i.e., from no core to a minimum core size) predicts more asymmetric structure. On the other hand, in Section 4.2, we focus on an increase in core size, conditional on positive measure of core size, which predicts more symmetric market structure and lower intermediation.

4.2.4 Normative Implications

Concentration Can be Efficient with Platform Access Thanks to Proposition 1, we know that efficiency is obtained if and only if the traders' private payoff is aligned with the social payoff. Thus, concentration itself might not be inefficient. Hence, the optimal intervention should not be targeting all-to-all trading or reducing risk concentration. Indeed, if there is no gap between private incentives of risk taking and entry cost, our results highlight the existence of exclusive core members and a high concentration of risks

[22]Bao, O'Hara, and Zhou (2016) and Bessembinder et al. (2018) show that the Volcker rule leads to lower inventories and capital commitment for bank-affiliated dealers. Such a decline, however, does not worsen the overall market liquidity measured by the bid-ask spread.

and volume can be efficient simply because of the economic scale, which is driven by the fixed cost component of entering platform.

Welfare-maximizing Policy with Entrenchment by Incumbent Cores On the other hand, whenever there are frictions that lead to a deviation between private incentives for risk taking and the entry cost, the equilibrium can be inefficient. According to Proposition 1, such an inefficiency (if it exists) can be corrected by aligning the private and social value of risk taking and/or entry.

One common concern, for example, is that the platform might be controlled by or have entrenched incumbent dealers. One can capture this in our environment by assuming that a set I_0 of agents with exogenous measure $\frac{c_0}{2^N}$ have built relationships among themselves and collectively operate the trading platform at cost ϕ. The incumbent agents jointly own the platform and decide whether to charge a new entrant to the platform an exogenous fee $\Delta > 0$.

Given any fee, this setup can thus be understood as our trading game with heterogeneous costs ϕ_i where $\phi_i \equiv \phi + \Delta$ for potential entrants $i \notin I_0$ and $\phi_i = \phi$ for incumbent banks $i \in I_0$. That is, the incumbent cores have a lower entry cost than the rest of the market. The existence of the fee thus generates the wedge between the private and social value of platform. Our model thus predicts that setting the subsidy s^c for entry such that $c^*(\phi + \Delta - s^c) = c^*(\phi)$ or introducing a new platform with entry cost ϕ will restore the efficient market structure.

5 Conclusions

In this paper, we develop a tractable framework of endogenous trading networks and use it to analyze how the market structure may respond to underlying parameters and/or regulatory changes. Exactly because banks can accumulate risks from others, any policy must account for the network effect of risk-taking behaviors among banks. Although the network structure seems complex, our framework provides a tractable and unique characterization as well as simple guidelines for possible interventions when private incentives are distorted relative to the social cost.

A Appendix

A.1 Efficiency, Uniqueness, and Variance Representation

Because agents have quasilinear utility, Pareto optimal allocations are the solution to a simple social planner's optimization problem where the planer maximizes the present value of total utility of the economy. The planner's choices in period t include any agent i's counterparty $j_{i,t}$, asset allocation within a match, $\tilde{a}_{i,t+1}(a_{i,t}, a_{j_{i,t},t})$ and $\tilde{a}_{j_{i,t},t+1}(a_{i,t}, a_{j_{i,t},t})$. The planner chooses period-t counterparties given period-0 information and asset distribution in period t. The planner's value function in period t has the joint asset distribution across agents as its state variable and can be characterized as

$$\Pi_t(\pi_t) = \int E_1 u_t(\widetilde{a}_{i,t}(a_{i,t}, a_{j_{i,t},t}))di + \beta \Pi_{t+1}(\pi_{t+1}), \text{ for } t \leq N,$$

$$\Pi_{N+1}(\pi_{N+1}) = \int E_1 u_{N+1}(a_{i,N+1})di.$$

The constraints that the planner faces include:

(1) Given π_t, the planner's period-t is feasible if and only if

$$\int_0^i \Pr(j_{\iota,t} \leq \iota)d\iota \leq i, \tag{A.1}$$

$$\tilde{a}_{i,t}(a_i, a_{j_{i,t}}) + \tilde{a}_j(a_i, a_{j_{i,t}}) = a_i + a_{j_{i,t}}, \tag{A.2}$$

where (A.1) is the feasibility constraint of the matching allocation of the planner, $\Delta(\pi_{i,t})$ refers to the support of the marginal distribution $\pi_{i,t}$; (2) The joint distribution evolves consistently with the counterparty assignment and within match asset allocations.

Proposition 1 holds because the equilibrium value of bank i in period t equals the shadow value of adding bank i to the planner's optimization problem in period t. For a more detailed proof, see for example Chang and Zhang (2022).

Under the risk preferences specified in Section 3.1, because agents' utility is quadratic in their asset holding, only the mean and variance of a distribution are relevant to their payoff. In general, we can represent the joint distribution by the means and variances of agents' asset holdings and covariances between their asset holdings. To do this, we first show that it is optimal to keep the means of individual asset holding at zero. We then show that it is optimal to match agents whose asset holdings are not correlated.

Lemma 4. *It is optimal to keep the means of individual asset holding at zero.*

Proof. Assumption (3) can be translated into controlled changes in the mean and variance of an agent's asset holding. Denote $E_t a_{i,t} = m_{i,t}$, $E_t(a_{i,t} - m_{i,t})^2 = v_{i,t}$ and $\rho_{i,j,t} = \frac{Cov(a_{i,t+1}, a_{j,t+1})}{\sqrt{v_{i,t+1} v_{j,t+1}}}$ for all i, j, and t. Because the utility function of the agent is quadratic, the marginal asset distribution for Agent i enter the social planner's objective through its expected value and variance. Let $\boldsymbol{m}_t = \{m_{i,t}\}_{\forall i}$, $\boldsymbol{v}_t = \{v_{i,t}\}_{\forall i}$, $\boldsymbol{\rho}_t = \{\rho_{i,j,t}\}_{\forall i,j}$. Then the period-$t$ state variable of the social planner can be summarized by $(\boldsymbol{m}_t, \boldsymbol{v}_t, \boldsymbol{\rho}_t)$.

The planner's objective function is then

$$\Pi_t(\boldsymbol{m}_t, \boldsymbol{v}_t, \boldsymbol{\rho}_t) = -\int \kappa_{i,t} \left(m_{i,t+1}^2 + v_{i,t+1}\right) di + \beta \Pi_{t+1}(\boldsymbol{m}_{t+1}, \boldsymbol{v}_{t+1}, \boldsymbol{\rho}_{t+1}), \text{ for } t \leq N, \tag{A.3}$$

where

$$\Pi_{N+1}(\boldsymbol{m}_{N+1}, \boldsymbol{v}_{N+1}, \boldsymbol{\rho}_{N+1}) = \int W_{N+1}(v_{i,N+1}) di. \tag{A.4}$$

The feasibility of within-match asset allocation between agent i and her counterparty j implies that $a_{i,t+1} + a_{j,t+1} = a_{i,t} + a_{j,t}$ for all $t \leq N$, which is translated into two separate constraints for the mean and the variance of asset allocation to Agents i and j

$$m_{i,t+1} + m_{j,t+1} = m_{i,t} + m_{j,t}, \tag{A.5}$$

$$v_{i,t+1} + v_{j,t+1} + 2\sqrt{v_{i,t+1} v_{j,t+1}} \rho_{i,j,t+1} = v_{i,t} + v_{j,t} + 2\sqrt{v_{i,t} v_{j,t}} \rho_{i,j,t}. \tag{A.6}$$

Note that the choice of expected asset holding is subject to a separate constraint, (A.5), from the choice of its variance, (A.6). The law of motion of asset holding variance and correlation does not depend on the expected asset holding.

The planner's optimization problem in period t can be summarized by the following

Lagrangian:

$$\mathcal{L}_t(\boldsymbol{m}_t, \boldsymbol{v}_t, \boldsymbol{\rho}_t) = -\int \kappa_{i,t}\left(m_{i,t+1}^2 + v_{i,t+1}\right) di + \beta\Pi_{t+1}(\boldsymbol{m}_{t+1}, \boldsymbol{v}_{t+1}, \boldsymbol{\rho}_{t+1}) \quad \text{(A.7)}$$
$$+ \int \lambda^m_{i,j_{i,t},t}\left(m_{i,t} - m_{i,t+1}\right) di$$
$$+ \int \lambda^v_{i,j_{i,t},t}(v_{i,t} + \sqrt{v_{i,t}v_{j_{i,t}t}}\rho_{i,j_{i,t},t} - v_{i,t+1} - \sqrt{v_{i,t+1}v_{j_{i,t+1},t+1}}\rho_{i,j_{i,t},t+1})di$$

for all $t \leq N$, where $\lambda^m_{i,j_{i,t},t}$ refers to the Lagrangian multiplier for constraint (A.5) for agent i and her counterparty $j_{i,t}$, and $\lambda^v_{i,j_{i,t},t}$ refers to the Lagrangian multiplier for constraint (A.6).

For period $N+1$, $\frac{\partial \Pi_{N+1}(\boldsymbol{m}_{N+1},\boldsymbol{v}_{N+1},\boldsymbol{\rho}_{N+1})}{\partial m_{i,N+1}} = 0 \geq \frac{\partial \Pi_{N+1}(\boldsymbol{m}_{N+1},\boldsymbol{v}_{N+1},\boldsymbol{\rho}_{N+1})}{\partial v_{i,N+1}}$ and $\frac{\partial \Pi_{N+1}(\boldsymbol{m}_{N+1},\boldsymbol{v}_{N+1},\boldsymbol{\rho}_{N+1})}{\partial \rho_{i,j,N+1}} = 0$ for all i, j.

Using mathematical deduction, we can then show that $\frac{\partial \Pi_t(\boldsymbol{m}_t,\boldsymbol{v}_t,\boldsymbol{\rho}_t)}{\partial m_{i,t}} \leq 0$ for all i and all $t \leq N$, where the inequality is strict if and only if there exits $t \leq t' \leq N$ such that $\kappa_{t'} > 0$. This is because given the counterparty choices, $j_{i,t}$, the first order condition with respect to $m_{i,t+1}$ implies that $\lambda^m_{i,j_{i,t},t} < 0$ when $\kappa_t > 0$ or $\frac{\partial \Pi_{t+1}(\boldsymbol{m}_{t+1},\boldsymbol{v}_{t+1},\boldsymbol{\rho}_{t+1})}{\partial m_{i,t+1}} < 0$.

The effect of within-match asset allocation on Agent i's expected asset holding can be summarized by $\alpha^m_{i,t}$, such that $m_{i,t+1} = \alpha^m_{i,t}(m_{i,t} + m_{j,t})$, $m_{j,t+1} = (1-\alpha^m_{i,t})(m_{i,t} + m_{j,t})$. If $\frac{\partial \Pi_{t+1}(\boldsymbol{m}_{t+1},\boldsymbol{v}_{t+1},\boldsymbol{\rho}_{t+1})}{\partial m_{i,t+1}} < 0$, it is clear that $\alpha^m_{i,t}$ should be between 0 and 1. If $\alpha^m_{i,t}$ were greater than 1 or less than 0, the planner can strictly increase either agent i or her counterparty $j_{i,t}$'s marginal contribution to the planner's period t objective function without reducing other agents' contribution. For example, if $\alpha^m_{i,t} > 1$, by setting $\alpha^m_{i,t}$ to 1 reduces $m^2_{i,t+1}$ to $(m_{i,t} + m_{j,t})^2$ and $m^2_{j_{i,t},t+1}$ to 0. If $\frac{\partial \Pi_t(\boldsymbol{m}_{t+1},\boldsymbol{v}_{t+1},\boldsymbol{\rho}_{t+1})}{\partial m_{i,t+1}} = 0$, but $\kappa_{i,t} > 0$, the same argument applies so that $0 \leq \alpha^m_{i,t} \leq 1$. If $\frac{\partial \Pi_t(\boldsymbol{m}_{t+1},\boldsymbol{v}_{t+1},\boldsymbol{\rho}_{t+1})}{\partial m_{i,t+1}} = 0$, and $\kappa_{i,t} = 0$, it is without loss to the social planner to impose $0 \leq \alpha^m_{i,t} \leq 1$.

Because the expected value of agents' initial marginal asset distribution is zero, the fact that $0 \leq \alpha^m_{i,t} \leq 1$ implies that $m_{i,t} = 0$ for all i and all period. □

Lemma 4 is the first step in characterizing the efficient asset allocation. It implies that the socially optimal asset distribution in any period can be represented by the variance of individual agents' asset holdings and the correlation of their asset holdings.

Lemma 5. *In the socially optimal matching assignments and asset allocations, the post trade asset holdings of two matched Agents i and j are perfectly correlated, and the planner* *always match agents with uncorrelated asset holding. That is,* $\rho_{i,j_{i,t},t} = 0$*, and* $\rho_{i,j_{i,t},t+1} = 1$*, for any agent* i *and their optimal counterparty* $j_{i,t}$*.*

Proof. The proof takes two steps. First, we show that if $\rho_{i,j_{i,t},t} = 0$ for any agent i and their optimal counterparty $j_{i,t}$, it is optimal to have within match asset allocation perfectly correlated.

If $\rho_{i,j_{i,t+1},t+1} = 0$, then for all i, j such that $\rho_{i,j,t+1} > 0$, we can show by differentiating the planner's Lagrangian, (A.7), that $\frac{\partial \Pi_{t+1}(\boldsymbol{m}_{t+1}, \boldsymbol{v}_{t+1}, \boldsymbol{\rho}_{t+1})}{\partial \rho_{i,j,t+1}} = 0$. Following a similar argument to that in the proof of Lemma 4, we can see that the marginal value of increasing an agent's variance is negative $\frac{\partial \Pi_{t+1}(\boldsymbol{m}_{t+1}, \boldsymbol{v}_{t+1}, \boldsymbol{\rho}_{t+1})}{\partial v_{i,t+1}} \leq 0$.

The feasibility of within-match asset allocation implies that variances of asset allocations satisfy (A.6). According to (A.6), increasing the correlation between asset allocations to matched agents reduces the total variance of asset allocation to them, $v_{i,t+1} + v_{j_{i,t},t+1}$. Because $\frac{\partial \Pi_{t+1}(\boldsymbol{m}_{t+1}, \boldsymbol{v}_{t+1}, \boldsymbol{\rho}_{t+1})}{\partial \rho_{i,j,t+1}} = 0$, it is then optimal to set $\rho_{i,j_{i,t},t+1} = 1$.

The second step is to show $\rho_{i,j_{i,t},t} = 0$. Because the initial asset holdings are not correlated, if $\rho_{i,j_{i,t},t+1} = 1$, then the asset allocations are either uncorrelated or perfectly positively correlated. Because there is a continuum of agents in the economy, for any agent i, if the planner is to match him with an agent with variance v', there always exists such an agent whose asset holdings are uncorrelated with agent i. According to (A.7), this shadow value of $\rho_{i,j_{i,t},t}$ equals $\lambda^v_{i,j_{i,t},t}$, which is weakly negative. It is then optimal to match two agents whose asset holdings are not correlated. □

Lemma 5 implies that even though agents have the option to trade repeatedly with a counterparty, repeated trade without receiving new asset holding shocks is suboptimal. Trading once, the asset holdings of Agent i and the counterparty become positively correlated. Then, trading twice is dominated by trading with a new counterparty with the same asset holding variance but whose asset holding is not correlated with Agent i's. Thus, we can characterize the equilibrium using a representation of the aggregate asset holding distribution by the variances of individual agents' asset holding distribution.

A.2 Network Properties

A.2.1 Proposition 2

For result (1): From Equation 11, let

$$f_t(\alpha) \equiv -\kappa_t \left\{\alpha^2 + (1-\alpha)^2\right\} V + W_{t+1}(\alpha^2 V) + W_{t+1}((1-\alpha)^2 V)$$

We thus have

$$f_t'(\alpha) = \left(-\kappa_t + W_{t+1}'(\alpha^2 V)\right) 2\alpha V - \left(-\kappa_t + W_{t+1}'((1-\alpha)^2 V)\right) 2(1-\alpha)V.$$

If $W_{t+1}'' < 0$,$F_t(\alpha)$ is a concave function in α, as

$$\begin{aligned} f_t''(\alpha) = &\left(-\kappa_t + W_{t+1}'(\alpha^2 V)\right) 2V + \left(-\kappa_t + W_{t+1}'((1-\alpha)^2 V)\right) 2V \\ &+ W_{t+1}''(\alpha^2 V)(2\alpha V)^2 + W_{t+1}''((1-\alpha)^2 V)\left(2(1-\alpha)V\right)^2 < 0. \end{aligned}$$

Hence, if $W_{N+1}(V)$ is concave in V, $\alpha = \frac{1}{2}$, which satisfies the FOC, is the global maximizer. Thus

$$\Omega_N(v_i + v_j) = -\kappa_N \left(\frac{v_i + v_j}{2}\right) + W_{N+1}(\frac{v_i}{2}) + W_{N+1}(\frac{v_j}{2}),$$

Given that $W_N(v_i) = \max_j \Omega_N(v_i + v_j) - W_N(v_j)$, we thus have

$$W_N'(v_i) = -\kappa_N + \frac{1}{2} W_{N+1}'(\frac{v_i}{2})$$

and hence, $W_N''(v_i) < 0$ if $W_{N+1}''(\frac{v_i}{2}) < 0$. By backward induction, we have $W_t''(v) < 0 \; \forall t, v$ and thus risk sharing is always the optimal solution. We thus have $v_{i,t} = \frac{1}{2} v_{i,t-1} = \left(\frac{1}{2}\right)^t v_0$ $\forall i$. Since all agents are symmetric over time, it is WLOG to assume random matching.

For result (2): Given that $V_{ij} = v_i + v_j$, to establish PAM, it is sufficient to show that

$\Omega_t(V)$ is convex in V $\forall t$. Let $\alpha = \alpha^*(V)$ denote the optimal allocation under V.

$$
\begin{aligned}
&\Omega_t(\lambda V) + \Omega_t((1-\lambda)V) \\
\geq &\kappa_t \left\{(\alpha^2 + (1-\alpha)^2)V\right\} + W_{t+1}(\alpha^2 \lambda V) + W_{t+1}((1-\alpha)^2 \lambda V) \\
&+ W_{t+1}(\alpha^2(1-\lambda)V) + W_{t+1}((1-\alpha)^2(1-\lambda)V) \\
\geq &\left\{\kappa_t(\alpha^2 + (1-\alpha)^2)V + W_{t+1}\left(\alpha^2\left(\lambda V + (1-\lambda)V\right)\right) + W_{t+1}\left((1-\alpha)^2\left(\lambda V + (1-\lambda)V\right)\right)\right\} = \Omega_t(V).
\end{aligned}
$$

where the first inequality follows that the surplus under optimal allocation $\alpha^*(\lambda V)$ and $\alpha^*((1-\lambda)V)$ is higher than using the allocation rule $\alpha^*(V)$. The second follows that $W_{t+1}(v)$ is convex in v, which is true for $W_{N+1}(v)$. Assume that $W_{t+1}(v)$ is convex, it thus implies that $\Omega_t(V_{ij})$ is convex in $V_{ij} = v_i + v_j$. Moreover, since

$$W_t(v_i) = \max_j \left\{\Omega_t(v_i + v_j) - W_t(v_j)\right\},$$

it thus shows that $W_t(v)$ is convex in v $\forall t$. Hence, by backward induction, $\Omega_t(v_i + v_j)$ is convex in $v_i + v_j$ and hence PAM $\forall t$.

A.2.2 Proof for Lemma 8

Proof. For any $\alpha(V)$ that satisfies the FOC condition and PAM, we thus have

$$\Omega_t(V|g_t) = \Sigma_k \left\{-\kappa_t \alpha_k^2 V + W_{t+1}(\alpha_k^2 V | g_{t+1}(\alpha_k^2 V))\right\},$$

where $\alpha_i = \alpha(V) = 1 - \alpha_j$.

By the envelope theorem, and $v = 2V$,$W_t(v|g_t) = \frac{1}{2}\Omega_t(2v|g_t)$, we have

$$
\begin{aligned}
W_t'(v|g_t) = \Omega_t'(2v|g_t) &= \left\{-\kappa_t + W_{t+1}'(\alpha^2 V|g_{t+1}(\alpha^2 V))\right\}\alpha^2 + \left\{-\kappa_t + W_{t+1}'((1-\alpha)^2 V|g_{t+1}((1-\alpha)^2 V))\right\}(1 \\
&= \frac{\prod_{k\in\{i,j\}}\left(-\kappa_t + W_{t+1}'(\alpha_k^2 V|g_{t+1}(\alpha_k^2 V))\right)}{\Sigma_{k\in\{i,j\}}\left(-\kappa_t + W_{t+1}'(\alpha_k^2 V|g_{t+1}(\alpha_k^2 V))\right)} = \frac{1}{2}H(-\kappa_t + W_{t+1}'(\alpha^2 V|g_{t+1}(\alpha^2 V)), -\kappa_t + W_{t+1}'((1-\alpha
\end{aligned}
$$

, where using the fact that from FOC $\alpha_k = \frac{-\kappa_t + W_{t+1}'(\alpha_{-k}^2 V|g_t(\alpha_{-k}^2 V))}{\Sigma_k\left(-\kappa_t + W_{t+1}'(\alpha_k^2 V|g_t(\alpha_k^2 V))\right)}$. □

A.2.3 Proof for Lemma ??

Let $F_N(\alpha) \equiv \frac{1}{2^N}\left\{W_{N+1}\left(\alpha^2\left(V\right)\right) + \left(2^N-1\right)W_{N+1}\left(\left(\frac{1-\alpha}{2^N-1}\right)^2 V\right)\right\}$,the FOC thus yields

$$F_N^{(1)}(\alpha|N) = 2\sqrt{V}\left\{W'\left(\left(\alpha^2 V\right)\right)\sqrt{\alpha^2 V} - W'\left(\left(\frac{1-\alpha}{2^N-1}\right)^2 V\right)\sqrt{\left(\frac{1-\alpha}{2^N-1}\right)^2 V}\right\}. \tag{A.8}$$

and $\chi(v) \equiv \left\{\left(\frac{dW'(v)\sqrt{v}}{dv}\right)\sqrt{v}\right\} = \frac{1}{2}W'(v) + W''(v)v$, SFOC can be rewritten as

$$F_N^{(2)}(\alpha) = 4V\left(\chi(\alpha^2 V) + \chi\left(\left(\frac{1-\alpha}{2^N-1}\right)^2 V\right)\frac{1}{2^N-1}\right), \tag{A.9}$$

which is concave as $\chi(v)$ is concave. Let $\bar{v}$ denote the maximum of $\chi(v)$. Under A3, $\chi(0) = \frac{1}{2}W'(0) < 0$ and $\lim_{v\to\infty}\chi(v) > 0$, and we thus have $\chi(\bar{v}) > 0$, and there exists $\hat{v} < \bar{v}$ such that $\chi(\hat{v}) = 0$, and $\chi(v) < 0$ iff $v < \hat{v}$. This also implies that $\frac{dW'(v)\sqrt{v}}{dv} = \frac{\chi(v)}{\sqrt{v}} < 0$ iff $v < \hat{v}$. In other words, $W'(v)\sqrt{v}$ is a unimodal function with the minimum at $\hat{v}$. Hence, for any asymmetric root that satisfies FOC , it must be the case that $v_l(\alpha) < \hat{v} < v_h(\alpha)$.We now use the next two lemma to establish that there can be at most one core.

Lemma 6. *When $N = 1$, $F_N(\alpha)$ is single-peaked. Moreover, the optimal share $\alpha_N^*(v)$ at period N continuously increases in v*

Proof. Since

$$F^{(3)}\left(\frac{1}{2^N}\right) = \left\{g'\left(\frac{1}{2^N}V\right)\sqrt{\frac{1}{2^N}V}\right\}\left(1 - \left(\frac{1}{2^N-1}\right)^2\right), \tag{A.10}$$

hence when $N = 1$, $F_N^{(3)}(\frac{1}{2}) = 0$. Given that $F_N^{(3)}(\frac{1}{2})$ is concave, $F_N^{(2)}(\frac{1}{2})$ is the maximum of $F_N^{(2)}(\alpha)$ and $F_N^{(3)}(\alpha) < 0\ \forall\alpha \in (\frac{1}{2}, 1)$. This thus means that there can be at most one local maximum in the region of $(\frac{1}{2}, 1)$. To see this, suppose that there are two maxima (α_1, α_2) in this region; then, there must exist a local minimum $\alpha_{min} \in (\alpha_1, \alpha_2)$, where $F_N^{(2)}(\alpha_{min}) > 0$ but $F_N^{(2)}(\alpha_1) < 0$ and $F^{(2)}(\alpha_1) < 0$, which again contradicts that $F^{(3)}(\alpha) < 0\ \forall\alpha \geq \frac{1}{2}$. Hence, (1) if $F^{(2)}(\frac{1}{2}) = g(\frac{V}{4}) < 0$, then $\frac{1}{2}$ is the unique local maximum; (2) if $F^{(2)}(\frac{1}{2}) = g(\frac{V}{4}) > 0$, then $\frac{1}{2}$ is the local minimum, and there is a unique local maximum $\alpha^* \in (\frac{1}{2}, 1)$.

By the implicit function theorem, for any $\alpha_N^*(V) \geq \frac{1}{2}$, since $\chi(v) = \frac{1}{2}W'(v) + W''(v)v$, and we have

$$\begin{aligned}\frac{d\alpha_N^*(V)}{dV} &= -\frac{\frac{\partial^2 F_N(\alpha,V)}{\partial\alpha\partial V}}{\frac{\partial^2 F_N(\alpha,V)}{\partial^2\alpha}} |_{\alpha=\alpha^*} \propto 2\left\{W''_{N+1}(\alpha^2 V)\alpha^3 V - W''_{N+1}((1-\alpha)^2 V)(1-\alpha)^3 V\right\} \\ &= 2\left\{\chi(v_h(\alpha)) - \frac{1}{2}W'(v_h(\alpha))\right\}\alpha - \left\{\chi(v_l(\alpha) - \frac{1}{2}W'(v_l(\alpha))\right\}(1-\alpha) \\ &= 2\left\{\chi(v_h(\alpha))\alpha - \chi(v_l(\alpha))(1-\alpha)\right\} \geq 0.\end{aligned}$$

The first equality uses the fact that $\chi(v) = \frac{1}{2}W'(v) + W''(v)v$, and the second uses $W'(v_h(\alpha))\sqrt{\alpha} = W'(v_h(\alpha))\sqrt{(1-\alpha)}$ at α^*. The last inequality uses the fact that, for any $\alpha_N^*(V) > \frac{1}{2}$, it must be the case that $v_l(\alpha) < \hat{v} < v_h(\alpha)$; thus $\chi(v_l(\alpha)) < \chi(v_h(\alpha))$. □

Lemma 7. *Under A3 and $\kappa_t = 0$, (1) there can be at most two different values of v_{N+1}; (2) if $\max\{v_{N+1}^k\} > \min\{v_{N+1}^k\}$, there can be at most one core agent when $\kappa_t = 0$.*

Proof. First, from Equation A.8, $F^{(1)}(1) = W'(V) < 0$, which means that the solution must be interior. For Result (1), observe that any v_{N+1} must satisfy the FOC from the static problem, where $\{v_{k,N+1}\}$ maximizes

$$\max \Sigma_{k=1}^{2^N} W_{N+1}(v_{k,N+1}) \tag{A.11}$$

$$\left[\Sigma_k \sqrt{v_{k,N+1}}\right]^2 = 2^N v_1. \tag{A.12}$$

Hence,

$$\sqrt{v_{k,N+1}}\left(W'_{N+1}(v_{k,N+1})\right) = \lambda\sqrt{2^N v_1}, \tag{A.13}$$

where λ is Lagrange multiplier of the constraint A.12. Since $W'(v)\sqrt{v}$ is a unimodal function with the minimum at $\hat{v}$, Hence, there can be at most two roots for Equation A.13.

For Result (2): let $v_{N+1}^c = \max\{v_{N+1}^k\}$ and $v_{N+1}^0 = \min\{v_{N+1}^k\}$. This statement holds automatically when $N = 1$. We now show this holds when $N \geq 2$. Suppose that there are more than one agent with v_{N+1}^c. Given that the outcome can be achieved under any ordering of matching, first consider the case that the core is matched with a non-core agent in period N, which means that $v_N^1 = \frac{(\sqrt{v_{N+1}^c} + \sqrt{v_{N+1}^0})^2}{2}$ and it must be the case that

$\alpha_N^*(v_N^1) > \frac{1}{2}$. The same outcome, however, can be achieved by have two core agents meet in period N,which implies their $v_N^2 = \frac{(\sqrt{v_{N+1}^c}+\sqrt{v_{N+1}^c})^2}{2}$ and they adopt risk sharing, where $\alpha_N^*(v_N^2) = \frac{1}{2}$. Since $v_N^1 < v_N^2$, the fact that $\alpha_N^*(v_N^2) = \frac{1}{2}$ but $\alpha_N^*(v_N^1) > \frac{1}{2}$ violates the fact that $\alpha_N^*(v_N)$ increases in v_N. Contradictions.

Lastly, given that there can be at most one core agent and the problem is identical to a static allocation, Equation 16 thus follows from Equation A.11, using the constraint that $\left[\sqrt{\alpha^2 V} + \left(2^N - 1\right)\sqrt{v_l(\alpha)}\right]^2 = 2^N v_1$, we thus have $v_l(\alpha) = \left(\frac{\sqrt{V}-\sqrt{\alpha^2 V}}{(2^N-1)}\right)^2$. □

A.2.4 Proof for Proposition 3

Proof. Step 1: We first show that, for any $N > 1$, $F_N(\alpha)$ has at least two local maxima (α_e^*, α_c^*) from some mid-range of $V \in (V_\ell, V_h)$, where $\alpha_e^* = \frac{v_1}{2^N}$ represents the risk sharing and $\alpha_c^* \in (\frac{v_1}{2^N}, 1)$ is a solution involve risk concentration. To do so, we show that $F_N(\alpha)$ is convex in some region (α_1, α_2), where $\frac{1}{2^N} < \alpha_1 < \alpha_2 < 1$.

First of all, in order to guarantee that full risk sharing is a local maximum, we need

$$F_N^{(2)}(\frac{1}{2^N}) = 4V\left\{g(\frac{V}{(2^N)^2})\frac{2^N}{2^N-1}\right\} < 0,$$

hence, this condition holds whenever $\left(\frac{1}{2^N}\right)^2 V < \hat{v}$. Hence, we set $V_h = 2^N v_1 = \left(2^N\right)^2 \hat{v}$.

Moreover, from Equation A.10,

$$F_N^{(3)}(\frac{1}{2^N}) = \left\{\chi'(\left(\frac{1}{2^N}\right)^2 V)\sqrt{\frac{1}{2^N}V}\right\}\left(1 - \left(\frac{1}{2^N-1}\right)^2\right) > 0,$$

as $\chi'(v) > 0$ for $v < \hat{v} < \bar{v}$. To show that $F_N^{(2)}(\alpha) > 0$ for some interior range of α, it is sufficient to show that the maximum value of $F_N^{(2)}(\alpha)$ is large than zero. Let $\Gamma(V) = \max_\alpha F_N''(\alpha|V)$. One can show that $\Gamma(v)$ increases in v. To see this, let $\hat{\alpha}(V)$ be the solution above. For any $V' > V$, let $\tilde{\alpha} = \hat{\alpha}(V)\sqrt{\frac{V}{V'}}$, and thus

$$\begin{aligned}\Gamma(V') &\geq \chi\left(\left(\frac{\hat{\alpha}^2(V)V}{V'}\right)V'\right) + \chi\left(\left(\frac{1-\left(\frac{\hat{\alpha}(V)V}{V'}\right)}{2^N-1}\right)^2 V'\right)\left(\frac{1}{2^N-1}\right) \\ &= \chi\left(\hat{\alpha}^2(V)V\right) + \chi\left(\left(\frac{\sqrt{V'}-\left(\hat{\alpha}(V)\sqrt{V}\right)}{2^N-1}\right)^2\right)\left(\frac{1}{2^N-1}\right) \\ &\geq \chi\left(\hat{\alpha}^2(V)V\right) + \chi\left(\left(\frac{\sqrt{V}-\left(\hat{\alpha}(V)\sqrt{V}\right)}{2^N-1}\right)^2\right)\left(\frac{1}{2^N-1}\right) = \Gamma(V),\end{aligned}$$

where the last inequality uses the fact that $\chi'(v_l(\alpha)) \geq 0$. Let $\bar{V}$ such that $\Gamma(\bar{V}) = 0$, we thus have $\Gamma(V) > 0$, for $V > \bar{V}$, and by continuity, there exists a region where $F''_N(\alpha|V) > 0$. Moreover, since

$$\Gamma(V_h + \epsilon) > \chi(\hat{v} + \frac{\epsilon}{2^N})\frac{2^N}{2^N-1} > 0,$$

it must be the case that to $\bar{V} < 2^N\hat{v}$. Lastly, we need to have

$$F''_N(1|V) = \chi(V) + \chi(0)\frac{1}{2^N-1} < 0, \tag{A.14}$$

so that $F''(\alpha|V)$ is concave when α is large enough. Note that since $F'_N(1|V) < 0$, together with $F''_N(1|V) < 0$, it then guarantees the existence of another local maximum $\alpha^*_c < 1$. Condition A.14 is possible as $2^N v_1 > \bar{v}$, $\chi'(v) < 0$ for $v < \bar{v}$ and $\chi(0) < 0$, this condition is thus guarantees when v_1 is large enough. Let $\tilde{V} > \bar{v}$ such that $F''_N(1|\tilde{V}) = 0$, we thus have $F''_n(1|V) < 0$ for $V > \tilde{V}$. Hence, set $V_\ell = \max\{\bar{V}, \tilde{V}\}$, there exists a region where $F''_N(\alpha|V) > 0$ when $V \in (V_\ell, V_h)$.

Step 2: We now show that the exists $V^* \in (V_\ell, V_h)$ such that α^*_e is the global optimal iff $V < V^*$.

$$D(V,N) \equiv \max_{\alpha > \frac{1}{2^N}}\left\{W_{N+1}\left(\alpha^2 V\right) + \left(2^N-1\right)W_{N+1}\left(\left(\frac{1-\alpha}{2^N-1}\right)^2 V\right)\right\} - 2^N W_{N+1}\left(\left(\frac{1}{2^N}\right)^2 V\right)$$

$$\frac{\partial D(V,N)}{\partial V} = \left\{ \left\{ W'_{N+1}\left(v_h(\alpha)\right)\frac{v_h(\alpha)}{V} + \left(2^N-1\right)W'_{N+1}\left(v_l(\alpha)\right)\frac{v_l(\alpha)}{V} \right\} - 2^N W'(\left(\frac{1}{2^N}\right)^2 V)\left(\frac{\left(\frac{1}{2^N}\right)^2 V}{V}\right) \right\}$$
$$= W'_{N+1}\left(v_l(\alpha)\right)\sqrt{v_l(\alpha)}\frac{1}{V}\left\{\sqrt{v_h(\alpha)} + \left(2^N-1\right)\sqrt{v_l(\alpha)}\right\} - W'(\left(\frac{1}{2^N}\right)^2 V)\frac{1}{2^N}$$
$$= W'_{N+1}\left(v_l(\alpha)\right)\sqrt{v_l(\alpha)} - W'(\left(\frac{1}{2^N}\right)^2 V)$$
$$= \frac{1}{\sqrt{V}}\left\{ W'_{N+1}\left(v_l(\alpha)\right)\sqrt{v_l(\alpha)} - W'(\left(\frac{1}{2^N}\right)^2 V)\sqrt{\frac{V}{\left(2^N\right)^2}} \right\} > 0$$

where the first equality uses FOC and thus $W'\left((\alpha^2 V)\right)\sqrt{\alpha^2 V} - W'\left(\left(\frac{1-\alpha}{2^N-1}\right)^2 V\right)\sqrt{\left(\frac{1-\alpha}{2^N-1}\right)^2 V} = 0$, the second equality uses the variance constraint $\sqrt{v_h(\alpha)} + \left(2^N-1\right)\sqrt{v_l(\alpha)} = \sqrt{V}$, and the last inequality uses the fact that $\frac{d\left(W'(v)\sqrt{v}\right)}{dv} < 0$ for $v < \hat{v}$. □

A.2.5 Proof for Corollary 4

We first show that, according to Lemma 8, when $\kappa_{t=0}$, we have $W'_t(v_{i,t}) = \frac{1}{2^{N-t+1}}\left\{\frac{2^{N-t+1}}{\Sigma_{k\in\Psi_t(i)}\left(\frac{1}{W'_{N+1}(v_{k,N+1})}\right)}\right\}$. This holds for period N. Assume that $W'_{t+1}(v_{i,t+1}) = \left\{\frac{1}{\Sigma_{k\in\Psi_t(i)}\left(\frac{1}{W'_{N+1}(v_{k,N+1})}\right)}\right\}$,by backward induction, we thus have

$$W'_t(v_{i,t}) = \frac{1}{2}\left\{\frac{2}{\Sigma\frac{1}{W'_{t+1}(v_{i,t+1})}}\right\} = \frac{1}{\Sigma_{k\in\Psi_{t+1}(i)}\left(\frac{1}{W'_{N+1}(v_{k,N+1})}\right) + \Sigma_{k\in\Psi_{t+1}(j)}\left(\frac{1}{W'_{N+1}(v_{k,N+1})}\right)}$$
$$= \frac{1}{\left\{\Sigma_{k\in\Psi_t(i)}\left(\frac{1}{W'_{N+1}(v_{k,N+1})}\right)\right\}^{-1}}$$

Hence, the value above only depends on whether $i_c \in \Psi_t(i)$. If so, $c_{i,t} = 1$, and thus have

$$\gamma_t(1) = \frac{1}{\frac{1}{W'_{N+1}(v^c_{N+1})} + \frac{(2^{N-t+1}-1)}{W'_{N+1}(v^0_{N+1})}} > \frac{1}{\frac{1}{W'_{N+1}(v^0_{N+1})} + \frac{(2^{N-t+1}-1)}{W'_{N+1}(v^0_{N+1})}} = \frac{W'_{N+1}(v^0_{N+1})}{2^{N-t+1}} = \gamma_t(0),$$

where the inequality uses the fact that $\frac{1}{W'_{N+1}(v^c_{N+1})} < \frac{1}{W'_{N+1}(v^0_{N+1})}$.

A.2.6 Proof for Lemma 3

We first use the lemma below to establish an additional necessary condition for the optimal path of $v_{i,t}$.

Lemma 8. *When the flow marginal cost of bearing risks* $\kappa_t > 0$, *the optimal solution must satisfy*

$$\tilde{v}^h_{t+1}(\tilde{v}^h_t(v_t)) \geq \tilde{v}^h_{t+1}(\tilde{v}^l_t(v_t)) \geq \tilde{v}^l_{t+1}(\tilde{v}^l_t(v_t)) \geq \tilde{v}^l_{t+1}(\tilde{v}^h_t(v_t)).$$

Proof. We now show that if this condition is violated, fixing $v_{k,t+2}$ but changing the ordering of the matches among these four banks lowers the total variance of $v_{k,t+1}$. Hence, whenever $\kappa_t > 0$, such a deviation is profitable. Given that the constraint yields $(\sqrt{v_{i,t+2}}+\sqrt{v_{j,t+2}})^2 = 2v_{i,t+1}$, we thus have,

$$\Omega_t(v) = -\kappa_t \frac{1}{2}\left\{\left[\sqrt{v_{i,t+2}} + \sqrt{v_{j,t+2}}\right]^2 + \left[\sqrt{v_{i,t+2}} + \sqrt{v_{j,t+2}}\right]^2\right\} + \Sigma_k\left(-\kappa_{t+1}v_{k,t+2} + W_{t+2}(v_{k,t+2})\right)$$
$$\leq -\kappa_t \frac{1}{2}\left\{\underbrace{\left[\sqrt{v_{1,t+2}} + \sqrt{v_{4,t+2}}\right]^2}_{v_{1,t+1}} + \underbrace{\left[\sqrt{v_{2,t+2}} + \sqrt{v_{3,t+2}}\right]^2}_{v_{2,t+1}}\right\} + \Sigma_k\left(-\kappa_{t+1}v_{k,t+2} + W_{t+2}(v_{k,t+2})\right)$$

The first inequality uses the fact that $f(v_i, v_j) \equiv \left[\sqrt{v_i} + \sqrt{v_j}\right]^2$ and $f_{12}(v_i, v_j) > 0$; hence NAM sorting minimizes the flow payoff. The last equality uses the fact that

$$v_t = \left(\sqrt{v_{1,t+1}} + \sqrt{v_{2,t+1}}\right)^2 = \frac{1}{4}\left[\sqrt{v_{1,t+2}} + \sqrt{v_{4,t+2}} + \sqrt{v_{2,t+2}} + \sqrt{v_{3,t+2}}\right]^2.$$

In other words, different matching plan in period $t+1$ only affects changes the flow payoff in period t. Hence, if the condition is violated, changing the ordering of the matches among these four banks lowers the total variance of $v_{k,t+1}$ and still have identical $v_{k,t+2}$ from period $t+2$ onward .

We now use the result above to prove Lemma 3.

Given the payoff in the final period $N+1$, $W_{N+1}(v) = \max_{c_{N+1}} \gamma_{N+1}(c_{N+1})v - \phi(c_{N+1})$, where $\gamma_{N+1}(c_{N+1})$ increase in $c_{N+1} \in \{0,1\}$, and thus if $c_{i,N+1} > c_{j,N+1}$, then it must be the case that $v_{i,N+1} > v_{j,N+1}$. Since $c_{i,N} = c_{i,N+1} + c_{jt(i),N+1} \in \{0,1,2\}$,the value of $\gamma_N(c)$ is given by Equation 14, which increases in c. Thus, $c^*_{N+1}(v)$ must increase in v.

For any period $t = N-1$, suppose that $c_{j,N} - c_{i,N} \geq 2$. which is only possible when

$c_{j,N} = 2$ and $c_{i,N} = 0$, as $c_{k,N} \in \{0, 1, 2\}$. Since $\gamma_N(c)$ increases in c, Agent j must hold strictly higher posttrade variance (i.e., $v_{j,N} > v_{i,N}$). Moreover, as $c_{k,N} \in \{0, 1, 2\}$, $c_{j,N} - c_{i,N} \geq 2$ is only possible when $c_{j,N} = 2$ and $c_{i,N} = 0$. This thus means that $c_{j,N+1} = c_{j_N(j),N+1} = 1$ and $c_{i,N+1} = c_{j_N(i),N+1} = 0$. Since $c^*_{N+1}(v)$ must increase in v, it thus implies that

$$\min\{v_{j,N+1}, v_{j_N(j),N+1}\} > \min\{v_{i,N+1}, v_{j_N(i),N+1}\},$$

which contracts Lemma 8. Hence, for any $c_{N-1} \in \{0, 1, 2, 3, 4\}$, the connections are unique, where $c_{i,N} = \left\{\lfloor \frac{c_{i,N-1}}{2} \rfloor, \lceil \frac{c_{i,N-1}}{2} \rceil\right\}$ and thus c_{N-1} is sufficient statics. Last, since $\gamma_N(c)$ decrease in c, $\gamma_{N-1}(c)$ thus also increases in c.

By backward induction, assume that $c_{i,t} = \left\{\lfloor \frac{c_{i,t}}{2} \rfloor, \lceil \frac{c_{i,t}}{2} \rceil\right\}$ and let $\gamma_{t+1}(c)$ denote its corresponding risk capacity, which decreases in c and the value function yields

$$W_t(v) = \max_c \gamma_t(c)v - \phi(c),$$

and hence if $c_{i,t} > c_{j,t}$, then it must be the case that $v_{i,t} > v_{j,t}$. Hence, by similar logic, if $c_{j,t+1} - c_{i,t+1} \geq 2$, then

$$\min\left\{c_{j,t+2}, c_{j^*_{t+1}(j),t+2}\right\} > \min\left\{c_{i,t+2}, c_{j^*_{t+1}(i),t+2}\right\}$$

and thus

$$\min\{v_{j,t+2}, v_{j^*_{t+1}(j),t+2}\} > \min\{v_{i,t+2}, v_{j^*_{t+1}(i),t+2}\},$$

which violates Lemma 8. Last, since $\gamma_{t+1}(c)$ is decreasing in c and, under optimal access, $\gamma_t(c) = \frac{1}{2}H(\kappa_t + \gamma_{t+1}(\lfloor \frac{c}{2} \rfloor), \kappa_t + \gamma_{t+1}(\lceil \frac{c}{2} \rceil))$ is thus increasing in c in period t. This thus establishes that Lemma 3 must hold for any t. □

A.2.7 Proof of Proposition 5

We first show that $\gamma^*_t(c|\delta, \eta, \kappa) = \kappa\gamma^*_t(c|\delta, \eta, 1)$ is a homogeneous function of κ. This holds for $N + 1$, as $\gamma_{N+1}(1) = -\eta\kappa$ and $\gamma_{N+1}(0) = -\kappa$. Given the expression for $\gamma^*_t(c|\delta, \eta, \kappa)$ from Equation 18, we thus have

$$\gamma_t^*(c|\delta,\eta,\kappa) = \frac{1}{2}H\left\{\kappa(-\delta+\gamma_{t+1}^*(\lfloor\frac{c}{2}\rfloor|\delta,\eta,1)),\kappa(-\delta+\gamma_{t+1}^*(\lceil\frac{c}{2}\rceil|\delta,\eta,1))\right\}$$
$$= \kappa\frac{1}{2}\left\{H\left(-\delta+\gamma_{t+1}^*(\lfloor\frac{c}{2}\rfloor|\delta,\eta,1)\right),\left(-\delta+\gamma_{t+1}^*(\lceil\frac{c}{2}\rceil|\delta,\eta,1)\right)\right\}.$$

Hence, Equation (19) can be rewritten as $\Pi = \kappa v_1 \max_c \left\{\hat{\gamma}_1(c) - \frac{c}{2^N}\left(\frac{\phi}{\kappa v_1}\right)\right\}$, where $\hat{\gamma}_1(c) = \gamma_t^*(c|\delta,\eta,1)$. By comparative statics, $c^*\left(\frac{\phi}{\kappa v_1}\right)$ increases in $\frac{\phi}{\kappa v_1}$.

A.3 Diminishing Marginal Cost of Bearing Risks and Endogenous Search Intensity: An Example

Suppose that banks can pay a quadratic cost $-\frac{c}{2}\gamma^2$ to have access to the competitive market with probability γ and that they choose the search intensity, γ, conditional their realized asset holding. Denote

$$W_{N+1}(v) = \int \widehat{W}(a)d\pi_{N+1}(a)$$

where $\widehat{W}(a)$ is a bank's expected payoff conditional on pretrade asset holding being a.

$$\widehat{W}(a) = \max_{\gamma} -\frac{c}{2}\gamma^2 - (1-\gamma)a^2$$

Thus, the optimal search intensity conditional on pretrade asset holding a is

$$\gamma(a) = c^{-1}a^2$$

and

$$\widehat{W}(a) = -\frac{c}{2}\gamma(a)^2 - (1-\gamma(a))a^2 = -a^2 + c^{-1}a^4.$$

If the asset holding follows a normal distribution with mean 0 and variance v, the kurtosis of the distribution is $Ea^4 = 3v^2$.

$$W_{N+1}(v) = E_v\widehat{W}(a) = -v + 3c^{-1}v^2$$

The marginal risk-bearing cost is decreasing in v, $W'(v) = -1 + 6c^{-1}v$.

References

Afonso, G. and R. Lagos (2015). "Trade dynamics in the market for federal funds". *Econometrica* 83.1, pp. 263–313.

Anderson, A. and L. Smith (2010). "Dynamic matching and evolving reputations". *The Review of Economic Studies* 77.1, pp. 3–29.

Babus, A. and T.-W. Hu (2017). "Endogenous intermediation in over-the-counter markets". *Journal of Financial Economics* 125.1, pp. 200–215.

Babus, A. and P. Kondor (2018). "Trading and Information Diffusion in Over-the-Counter Markets".

Babus, A. and C. Parlatore (2017). "Strategic fragmented markets". *Available at SSRN 2856629*.

Bao, J., M. O'Hara, and X. A. Zhou (2016). "The Volcker rule and market-making in times of stress". *Journal of Financial Economics, Forthcoming*.

Bessembinder, H. et al. (2018). "Capital commitment and illiquidity in corporate bonds". *The Journal of Finance* 73.4, pp. 1615–1661.

Cabrales, A., P. Gottardi, and F. Vega-Redondo (2017). "Risk sharing and contagion in networks". *The Review of Financial Studies* 30.9, pp. 3086–3127.

Chang, B. and S. Zhang (2018). "Endogenous market making and network formation". *Available at SSRN 2600242*.

Choi, J., Y. Huh, and S. Seunghun Shin (2023). "Customer liquidity provision: Implications for corporate bond transaction costs". *Management Science*.

Collin-Dufresne, P., B. Junge, and A. B. Trolle (2018). "Market structure and transaction costs of index CDSs". *Swiss Finance Institute Research Paper* 18-40.

Corbae, D., T. Temzelides, and R. Wright (2003). "Directed matching and monetary exchange". *Econometrica* 71.3, pp. 731–756.

Duffie, D. (2018). "Post-crisis bank regulations and financial market liquidity". *Lecture, Baffi*.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005). "Over-the-Counter Markets". *Econometrica* 73.6, pp. 1815–1847.

Dugast, J., S. Üslü, and P.-O. Weill (2019). *A Theory of Participation in OTC and Centralized Markets*. Tech. rep. National Bureau of Economic Research.

Eisfeldt, A. L., B. Herskovic, and S. Liu (2022). "Interdealer Price Dispersion". *Available at SSRN.*

Farboodi, M. (2014). "Intermediation and voluntary exposure to counterparty risk". *Available at SSRN 2535900.*

Gale, D. M. and S. Kariv (2007). "Financial networks". *American Economic Review*, pp. 99–103.

Glode, V. and C. C. Opp (2019). "Over-the-Counter versus Limit-Order Markets: The Role of Traders' Expertise". *The Review of Financial Studies.*

Gofman, M. (2011). "A network-based analysis of over-the-counter markets". In: *AFA 2012 Chicago Meetings Paper.*

Hojman, D. A. and A. Szeidl (2008). "Core and periphery in networks". *Journal of Economic Theory* 139.1, pp. 295–309.

Jackson, M. O. (2005). "A survey of network formation models: stability and efficiency". *Group Formation in Economics: Networks, Clubs, and Coalitions*, pp. 11–49.

Malamud, S. and M. Rostek (2014). "Decentralized Exchange".

McAfee, R. P. (1993). "Mechanism design by competing sellers". *Econometrica: Journal of the econometric society*, pp. 1281–1312.

Üslü, S. (2019). "Pricing and liquidity in decentralized asset markets". *Econometrica* 87.6, pp. 2079–2140.

Wang, C. (2016). "Core-periphery trading networks".

Yellen, J. (2013). "Interconnectedness and systemic risk: Lessons from the financial crisis and policy implications". *Board of Governors of the Federal Reserve System, Washington, DC.*

Yoon, J. H. (2017). *Endogenous market structure: Over-the-counter versus exchange trading.* Tech. rep. Working Paper, University of Wisconsin-Madison.