American Economic Review 2017, 107(11): 3320–3362
https://doi.org/10.1257/aer.20140759

Decentralized Exchange[†]

By Semyon Malamud and Marzena Rostek*

Most assets are traded in multiple interconnected trading venues. This paper develops an equilibrium model of decentralized markets that accommodates general market structures with coexisting exchanges. Decentralized markets can allocate risk among traders with different risk preferences more efficiently, thus realizing gains from trade that cannot be reproduced in centralized markets. Market decentralization always increases price impact. Yet, markets in which assets are traded in multiple exchanges, whether they are disjoint or intermediated, can give higher welfare than the centralized market with the same traders and assets. In decentralized markets, demand substitutability across assets is endogenous and heterogeneous among traders. (*JEL* D43, D44, D85, G11, G12)

In classical economic theory, markets are centralized. All units are exchanged through a single market clearing at the terms of trade that apply to all agents equally. In today's markets, essentially all financial assets are traded in multiple coexisting and interconnected trading venues. Trade away from centralized exchanges is common not only for assets and goods with heterogeneous units such as real estate, but also for homogeneous assets. Indeed, most bonds (government, municipal, and corporate) are traded over the counter, as are currencies, loans, and (more recently)

*Malamud: EPF Lausanne, Swiss Finance Institute, Quartier UNIL-Dorigny, Extranef 213 CH–1015 Lausanne, Switzerland, and CEPR (email: semyon.malamud@epfl.ch); Rostek: Department of Economics, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706 (email: mrostek@ssc.wisc.edu).
This paper was accepted under the guidance of Andrzej Skrzypacz, Coeditor. We would like to thank four anonymous referees for many helpful comments and suggestions. We also thank Ana Babus, Alberto Bisin, Bruce Carlin, Darrell Duffie, Matt Elliott, Mikhail Golosov, Sanjeev Goyal, Piero Gottardi, Christopher Henessy, Peter Kondor, Guido Lorenzoni, Anna Pavlova, Mark Ready, Yuliy Sannikov, Fernando Vega-Redondo, Rakesh Vohra, Nathan Yoder, Ji Hee Yoon, the participants at the 2013 AEA meetings (San Diego), the 9th Annual Cowles Conference on General Equilibrium and Applications (Yale), the Economic Networks and Banking Conference (LSE, 2013), SAET Conference (2013), SITE (2013), CEPR Asset Pricing Meeting (Gerzensee, 2013), Financial and Economic Networks Conference (Wisconsin School of Business, 2013), Cambridge-INET Workshop (2014), Conference in Honor of Pete Kyle (Maryland, Robert H. Smith School of Business, 2016) and the seminar participants at Academia Sinica, Bank of Canada, Berkeley, Boston College, Brown, Caltech, Carnegie Mellon/Pittsburgh, Collegio Carlo Alberto, CUHK, Duke, EUI, Fudan, Georgia State, Imperial College London, Iowa, Kellogg, LBS, Melbourne, Minnesota, MIT Sloan, Penn, Penn State, Purdue, Queensland, Rochester, SHUFE, UC Davis, UCL, UCLA, UNSW, UTS, and Yale. The authors gratefully acknowledge the financial support of this research by the NET Institute. Rostek acknowledges the financial support of the NSF through Grant SES-1357758. Malamud acknowledges the financial support of the Swiss Finance Institute and the Swiss National Science Foundation. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†]Go to https://doi.org/10.1257/aer.20140759 to visit the article page for additional materials and author disclosure statement(s).

stocks.[1] In fact, the past two decades have seen new types of marketplaces that offer different types of market clearing—direct matching with an intermediary, trading in a dealer network, or an electronic centralized exchange—to institutional and retail investors. This paper examines the potential for a decentralized market to create gains from trade. What are the economic mechanisms in decentralized markets that have no centralized market counterparts?

The growing literature on decentralized trading emphasizes important frictions that are associated with decentralization of trade, such as search, counterparty risk, or asymmetric information. With "decentralization" introduced as a friction in a competitive model, a typical result argues that the absence of frictions would correspond to maximal welfare. To understand the potential for welfare gains with decentralized trading, we consider markets with any number of strategic traders and divisible assets. The sole assumption of the centralized market model that we relax is that a single market clearing determines all agents' allocations. Namely, the *market* consists of *exchanges*, each defined by the subset of agents who trade there and the subset of assets traded, each with a separate market-clearing price. A market is *centralized* if there is a single exchange for all traders and assets and *decentralized* otherwise. Traders can participate in many exchanges in which the same or different assets may be traded. This accommodates market structures with coexisting exchanges, including centralized markets and empirically common market structures with private exchanges (with restricted participation), public exchanges, and intermediation. Preferences and assets are described by constant absolute risk aversion (CARA) utilities and Gaussian payoffs. Gains from trade come from risk sharing: endowments (which are agents' private information and are independent) and risk preferences are heterogeneous. Each exchange operates as a (uniform-price) double auction, and agents submit demand and supply schedules in the exchanges in which they participate. Thus, the model is a decentralized market counterpart of double auction models in the tradition of Kyle (1989), Vives (2011), and the capital asset pricing model (CAPM). This permits a direct comparison of predictions for centralized and decentralized markets.

Why might a decentralized market be more efficient? In markets with strategic traders, the Pareto efficiency result of the First Welfare Theorem does not apply: even if the total number of traders is large, traders in decentralized markets generally have a strictly positive price impact in the exchanges where they participate. In any market, centralized or decentralized, equilibrium allocations of each trader is a combination of his initial endowment and an *aggregate risk* portfolio (which corresponds to risk nondiversifiable in equilibrium). Our starting observation is that in noncompetitive centralized markets, unless all traders' risk preferences are symmetric, the aggregate risk portfolio that all traders trade toward differs from the efficient portfolio (which maximizes the total utility over feasible allocations, is independent

[1] In the US equity market, the NYSE executes less than one-quarter of the volume in its listed stocks; the remaining volume is created in over 10 public exchanges, more than 30 private exchanges (liquidity pools), and over 200 broker-dealer networks. Over the past few years, trading in private exchanges has grown by more than 50 percent in the US and has more than doubled in Europe (Schapiro 2010). Similarly, while prior to 2007, equity markets in Europe were characterized by dominant exchanges in each domestic market, the Markets in Financial Instruments Directive (MiFID) in 2007 created more than 200 new trading venues in which equities, bonds, and even derivatives are traded. As of 2012, these alternative venues accounted for at least 30 percent of total equity turnover. Duffie (2012) provides an overview.

of the market structure, and corresponds to risk nondiversifiable in the market, or *systematic risk*). In addition, the centralized market allocates risk in a particular way: relatively less risk-averse agents face a more inelastic residual supply and hence have larger price impact. Thus, less risk-averse agents will be reluctant to trade, and highly risk-averse traders will retain a large fraction of their endowment risk in equilibrium. This suggests that the efficiency of the centralized market allocation depends on the distribution of endowments among agents with different risk preferences. This further suggests that, to increase welfare, a decentralized market would have to *reallocate* systematic risk toward less risk-averse agents. We show that a decentralized market can increase the total welfare compared to the centralized market. In fact, simply breaking up the centralized market to create disjoint exchanges can increase welfare.

We present the following main results. First, when a market becomes more decentralized—some agents trade with fewer other traders or trade fewer assets—traders' price impacts weakly increase in any exchange and are thus lowest in the centralized market. This holds regardless of the asset structure in the more decentralized market. Furthermore, a general complementarity holds: any change in the market structure that lowers price impact "locally" in an exchange lowers price impact in all other exchanges as long as they are indirectly connected. For instance, creating new private exchanges, in which participation is restricted, weakly improves liquidity in the market.[2]

Second, although the centralized market minimizes price impact for all traders, a decentralized market with the same traders and assets may give higher total welfare. The key is that decentralized trading changes the agents' ability to diversify: since their participation in the exchanges differs, agents trade distinct components of the aggregate risk portfolio. Given that the centralized market aggregate risk portfolio is generally inefficient, utility can increase despite the higher price impact. Moreover, traders' equilibrium price impacts are no longer linked to risk preferences in the particular way in which they are with centralized trading. Essentially, by allowing heterogeneous access to traders and assets, a decentralized market can allocate risk among traders whose risk preferences differ more efficiently.

Our results imply that restricting trader participation, while increasing all traders' price impacts, may increase welfare. In particular, under conditions, the fact that trading is noncompetitive makes the case for various forms of market decentralization recently implemented or debated: breaking up exchanges and asset deconsolidation (such as ring-fencing of investment banking units and swaps push-out required by the Dodd-Frank Act and MiFID, demerger, or specialization in trading certain assets) may increase welfare;[3] moving an asset from centralized to OTC clearing and intermediation may increase welfare.

[2] One expects that this holds more generally when new exchanges do not affect too much inference about values in the existing exchanges. The creation of new exchanges has increased competition and substantially decreased liquidity costs in the US stock market. Similar decreases in trading costs occurred in Canada, Europe, and Asia, where different regulatory environments allowed electronic exchanges to develop earlier than those in the United States (Knight Capital Group 2010; Angel, Harris, and Spatt 2011; and O'Hara and Ye 2011).

[3] In practice, markets are decentralized (in our sense) also because different participants trade different assets—by choice or regulation. Even large financial institutions typically participate in only a few trading venues and trade a small subset of existing securities: e.g., pension funds cannot trade many types of derivatives, banks are allowed to hold but not trade loans, and most hedge funds have a clear specialization in trading a limited number of securities.

When should one expect a decentralized market to increase efficiency? We show that when all traders' risk aversions are symmetric, the centralized market maximizes welfare among all market structures. This holds regardless of the distribution of initial endowment risk. With sufficiently heterogeneous risk preferences, the welfare-maximizing market structure is decentralized for some endowment distributions, particularly those for which the lowest risk aversion agents hold large nondiversified endowments initially. More precisely, what matters is not the endowment shocks per se but how the inefficient part of endowments, and hence the need to trade, is distributed across traders. Welfare gains from decentralization exist even if the number of traders is arbitrarily large. Sufficient heterogeneity in risk preference needs to hold only for a few market participants, and endowments need not become more extreme as the market grows.

Underlying certain decentralized market effects that have no centralized market counterparts is that when trading is decentralized, demand substitutability for assets (i.e., demand Jacobian) is endogenous and generally differs among traders. Namely, we show that in centralized markets, all traders' equilibrium price impacts are always *proportional* to the assets' fundamental covariance. The factor of proportionality depends on the trader's risk aversion, with less risk-averse traders having greater price impact. In contrast, when the market is decentralized, the within- and across-exchange price impact induced by others' behavior depends on who participates in each exchange and which assets are traded so that price impact is generally not proportional to fundamental risk. This nonproportionality of incentives in risk implies that demand substitutability for the same assets is heterogeneous among traders who participate in different exchanges. In contrast, in centralized markets, agents' demand substitutability always corresponds to the assets' fundamental payoff substitutability and hence is the same for all traders. The endogenous demand substitutability creates incentives for agents to specialize in trading different assets—as specialist intermediaries or non-intermediating dealers—and may increase welfare in the Pareto sense, even without affecting aggregate risk. Specifically, by changing who trades which assets without necessarily changing who trades with whom, decentralizing a market may allow for a reduction in idiosyncratic risk, whose changes do not affect the risk that is nondiversifiable in equilibrium.

The methods we introduce will be useful to other researchers studying games and general equilibrium on networks and hypergraphs. A hypergraph generalizes a graph by allowing an edge to connect any number of nodes, beyond just two (e.g., Berge 1973). Relative to the existing literature on games on networks, such games allow applications with strategies that have multiple dimensions over actions and attributes. The model allows equilibrium analysis of traditional industrial organization questions (market power, competition, pricing, product and market design, mergers) in decentralized markets. Let us note that decentralized market games are not (super- or sub-)modular, and results that have been established for modular games do not allow us to draw conclusions about welfare in decentralized markets. We characterize the comparative statics of equilibrium and welfare with respect to preferences, assets, and market structure for decentralized markets with arbitrary market structures, multiple assets, and any number of strategic agents.

Related Literature.—This paper is part of the growing literature on decentralized markets. Most modern models are based on graphs, random or fixed. The random search and matching approach assumes that trade occurs in large markets among a continuum of traders, in which centralized trading would be efficient (e.g., Gale 1986a, b; Duffie, Gârleanu, and Pedersen 2005; Vayanos and Weill 2008; Weill 2008; Duffie, Malamud, and Manso 2009, 2014; Lagos and Rocheteau 2009; Lagos, Rocheteau, and Weill 2011; Afonso and Lagos 2012; Hugonnier, Lester, and Weill 2014; Atkeson, Eisfeldt, and Weill 2015). Empirically, while some markets are best described by random meetings among traders who are small relative to the market, in others (e.g., dealer networks or interbank systems) relationships are not random[4] and are dominated by large institutional investors, who have price impact. Dealing with price impact often serves as a primary motivation to create an OTC exchange.[5] Thus, the fact that agents are nonnegligible in trading matters for why markets are decentralized. This paper considers markets with any number of traders, all of whom are strategic. Allowing (not assuming) noncompetitive behavior is central to our predictions and turns out to be important for thinking about certain implications of decentralized trading. We take the market structure (who trades with whom) as given and, in this sense, are closer to the strand of literature that views agents as interacting on a fixed network (e.g., Kranton and Minehart 2001; Gale and Kariv 2007; Blume et al. 2009; Gofman 2011; Manea 2011; Nava 2015; Condorelli, Galeotti, and Renou 2017; Choi, Galeotti, and Goyal 2017; Babus and Kondor 2016; Elliott 2015; Rahi and Zigrand 2013; Bramoullé, Kranton, and D'Amours 2014). Like the random matching models, the existing networks literature largely views decentralization as a restriction on the efficiency of trade.

In addition, just as the standard equilibrium theory is based on a single market clearing for all traders and assets, the existing literature on decentralized trading makes the opposite assumption that all transactions are bilateral.[6] We study networked markets with coexisting exchanges (i.e., hypergraphs). A model which accommodates market structures described by hypergraphs is not essential to our welfare results. It allows studying equilibrium in any market structures "between" centralized and bilateral trading in which groups of traders interact. During the past two decades, transactions occurring outside open exchanges have largely shifted to "auctions" (market clearing via aggregation of multiple demands and supplies) introduced for large institutional traders, dealers and retail investors. Moving beyond bilateral links also enables one to examine *when* a decentralized market can behave like the centralized market, why, and the conditions under which equilibrium behavior differs.

[4] It is well documented that dealers or brokers trade via an established network structure and that trading relationships exist between banks. An average bank trades with a small number of counterparties, and most banks form stable relationships with at least one lending counterparty; e.g., the US Federal Funds market (Bech and Atalay 2010; Afonso, Kovner, and Schoar 2013), interbank markets (Craig and von Peter 2014; Cocco, Gomes, and Martins 2009), and US municipal bonds market (Li and Schürhoff 2012).

[5] For example, Biais, Bisière, and Spatt (2010); Knight Capital Group (2010); Angel, Harris, and Spatt (2011). Biais and Green (2007) attribute the historical shift in the US municipal and corporate bond markets toward OTC trading to the increased importance of institutional investors.

[6] Corominas-Bosch (2004) and Elliott (2011) allow for multilateral bargaining with search. Some models (see Duffie, Gârleanu, and Pedersen 2005; Lagos and Rocheteau 2009; and Lagos, Rocheteau, and Weill 2011) assume that trade can only happen through special intermediaries (dealers) who provide liquidity. Rahi and Zigrand (2013) study trade of price-taking investors intermediated by arbitrageurs.

Finally, both random and fixed graph models typically derive the terms of trade from bargaining (e.g., take-it-or-leave-it offers) or posted prices and have efficient surplus sharing on each link. In our model, trade occurs through the uniform-price double auction in which agents submit demand or supply schedules (equivalently and in practice, combinations of limit and market orders) in the exchanges where they participate. The uniform-price market clearing is a precise analog of the market clearing in centralized market models—in general equilibrium and their game-theoretic counterpart of games in demand and supply functions. As we show, traders have nonnegligible price impact—the behavior of prices and allocations differs from that in trading environments with efficient risk sharing per link.

I. A Decentralized Market Model

Market: Traders, Assets, and Exchanges.—Consider a market with I traders who trade K risky assets in N exchanges. Each exchange has separate clearing prices. We index agents by i, assets by k, and exchanges by n. An *exchange* $n \in N$ is identified by the subset of agents $I(n) \subseteq I$ who trade there and the subset of assets traded $K(n) \subseteq K$. The set of exchanges $\mathcal{M} = \{(I(n), K(n))\}_n$, which we take as a primitive, represents the *market structure*. Thus, $\mathcal{M}$ is a nonempty subset of the power set of $I \times K$, which, together with the set of agents I and assets K, corresponds to a hypergraph.[7] Agents can participate in many different types of trading venues for possibly nondisjoint subsets of traders (e.g., a public exchange, in which all traders participate; a private exchange, which restricts participation to a subset of traders; and intermediation). We assume that at least three agents participate in every exchange: $I(n) > 2$ for all n.[8]

The K risky assets have jointly normally distributed payoffs $R \sim \mathcal{N}(d, \Sigma)$ with positive definite covariance Σ; a riskless asset with a zero interest rate (a numéraire) is also available. Each trader i maximizes the expected CARA utility function $E[-\exp(-\alpha_i(-q_i^T p + (q_i^0 + q_i)^T R))]$, where α_i is agent i's absolute risk aversion, q_i^0 is his endowment vector of risky assets, q_i is i's vector of trades of risky assets, and p denotes the vector of prices. Endowments are (independent) private information. Using the fact that asset payoffs R_i are normally distributed, we have $E[-\exp(-\alpha_i(-q_i^T p + (q_i^0 + q_i)^T R))] = -\exp(-\alpha_i U_i(q_i))$ with

$$U_i(q_i) = d^T(q_i^0 + q_i) - \frac{\alpha_i}{2}\left(q_i^0 + q_i\right)^T \Sigma\left(q_i^0 + q_i\right) - p^T q_i, \tag{1}$$

and hence, equivalently, trader i maximizes the quasilinear-quadratic utility function (1).

In the analysis, *we treat assets traded in different exchanges as different assets*. That is, we do not impose a priori that identical assets (in the sense of $\mathcal{N}(d, \Sigma)$)

[7] A hypergraph is defined as a pair (X, E), where X is a set of elements called *nodes* and E is a set of nonempty subsets of X called (*hyper-*)*edges*. In our model, $X = (I, K)$ and an edge $(I(n), K(n))$ represents exchange n with $I(n)$ agents and $K(n)$ assets.

[8] As is well known, in centralized markets with two traders, a linear equilibrium with trade does not exist with independent private values (e.g., Kyle 1989). With negatively correlated values, equilibrium exists for any number of traders (Rostek and Weretka 2015b).

will trade at the same prices in different exchanges.[9] Thus, we treat a market with K assets traded in N exchanges $\{(I(n), K(n))\}_n$ as a market with $\sum_n K(n)$ (replicas of) assets and a $(\sum_n K(n)) \times (\sum_n K(n))$ positive semidefinite covariance matrix $\mathbf{V}$, induced by covariance Σ and the set of exchanges. This paper studies how the fact that exchanges are interlinked, through traders or assets, affects market behavior. Covariace matrix $\mathbf{V}$ describes the interconnectedness among the exchanges via traders and assets. We use capital bold notation for objects defined directly in a decentralized market.

A trader i who participates in a subset of all exchanges $N(i) \subseteq N$ maximizes

$$U_i(q_i) = d_{N(i)}^T(q_i^0 + q_i) - \frac{\alpha_i}{2}\left(q_i^0 + q_i\right)^T \mathbf{V}_{N(i)}\left(q_i^0 + q_i\right) - (p_{N(i)})^T q_i, \tag{2}$$

where $\mathbf{V}_{N(i)}$ is the submatrix of the covariance matrix $\mathbf{V}$, which corresponds to the assets traded by agent i in exchanges $N(i)$, $d_{N(i)}$ is the subvector of the expected payoff vector d, $q_i^0 \in \mathbb{R}^{N(i)}$ and $q_i \in \mathbb{R}^{N(i)}$ have the dimension of $\mathbb{R}^{N(i)}$ given by $\sum_{n \in N(i)} K(n)$,[10] and $p_{N(i)} = (p_j)_{j \in N(i)}$ denotes the vector of prices in exchanges $N(i)$. Market structure $\{(I(n), K(n))\}_n$ is equivalently described by trader *participation* in the exchanges $\{N(i)\}_i$, given agents I and assets K.

Example 1 (Centralized and Decentralized Markets):

(i) *The centralized market with multiple assets:* All agents I trade assets K in one exchange; $N = \{(I, K)\}$, and $\mathbf{V} = \Sigma$.

(ii) *A decentralized market for one asset:* All agents I trade the same asset 1 with a variance $\Sigma = \sigma^2$; $N = \cup\{(I(\ell), 1)\}_\ell$; and the covariance matrix $\mathbf{V}$ is singular and has rank 1, $\mathbf{V} = \sigma^2 \mathbf{1}$ where $\mathbf{1}$ is a matrix with all elements equal to 1.

(iii) *Private exchanges:* In addition to the *public exchange* for $K' \subseteq K$ assets, defined analogously to (i), there are L private exchanges (e.g., liquidity pools) in which only subsets of agents can trade different subsets of assets. There are L (possibly nondisjoint) subsets $I(1), \ldots, I(L) \subset I$ of agents, each trading in exchange $l \in L$; $N = \{(I, K')\} \cup \{(I(\ell), K(\ell))\}_\ell$ with $K' + \sum_\ell K(\ell)$ assets. Essentially all existing financial assets—most notably currencies, fixed income instruments (e.g., government and corporate bonds), and all derivatives—are traded in market structures that can be described as multiple interconnected private exchanges.

Agents can participate in many different types of trading venues for possibly non-disjoint subsets of traders (e.g., a centralized exchange, a dealer network,

[9] For a discussion of arbitrage, see footnote 21.

[10] We will be referring to the risky part of endowments, because this is the part that matters for the results. With K assets, the actual dimension of the endowment vector is K. We use notation $\mathbb{R}^{N(i)}$ to represent that the same asset can be traded in many exchanges: replicas of the same asset traded in different exchanges correspond to different coordinates, and the split of a given asset's endowments among the exchanges where it is traded is arbitrary. Equilibrium allocations do not depend on the split of the endowment across exchanges.

liquidity pools); there can be intermediation between traders in different types of trading venues.

Given the quasilinearity of the utility functions in cash (the numéraire), total welfare can be compared across market structures through the sum of the utilities (2) evaluated at the corresponding equilibrium allocations. Quasilinearity and market clearing imply that total welfare does not depend on prices,

$$\sum_i U_i(q_i) = \sum_i d^T q_i^0 - \sum_i \frac{\alpha_i}{2}\left(q_i^0 + q_i\right)^T \mathbf{V}_{N(i)}\left(q_i^0 + q_i\right). \tag{3}$$

It is clear that cash transfers exist for which one market Pareto dominates another if and only if the sum of equilibrium utilities is larger in the former.

Decentralized Double Auction.—Each exchange n operates as the standard uniform-price double auction for traders $i \in I(n)$ (e.g., Kyle 1989; Vives 2011; CAPM). Trader i submits a (net) demand schedule $q_i(p_{N(i)}) : \mathbb{R}^{N(i)} \to \mathbb{R}^{N(i)}$, which specifies demanded quantities of assets in the exchanges in which he participates; the demand is strictly downward-sloping in each exchange. The market clears simultaneously in all exchanges as $\sum_i \bar{q}_i(p^*_{N(i)}) = 0$, where $\bar{q}_i(p^*_{N(i)}) : \mathbb{R}^N \to \mathbb{R}^N$ is defined as equal to q_i on $N(i)$ and 0 on other exchanges $N \backslash N(i)$. This market-clearing condition defines the equilibrium price vector p^*; trader i receives $q_i \equiv q_i(p^*_{N(i)})$ and pays $p^*_{N(i)} \cdot q_i = p^* \cdot \bar{q}_i$. All traders are strategic; in particular, there are no noise traders.

Equilibrium.—We study the Bayesian Nash equilibrium in linear bid schedules (hereafter, *equilibrium*). With divisible goods, equilibrium is invariant to the distribution of independent private uncertainty.[11] Equilibrium with strategic traders in any market structure, centralized or decentralized, can be characterized through two conditions which correspond to individual optimization and market clearing (Proposition 1 and Theorem 1).

II. Equilibrium in Centralized versus Decentralized Markets

In this section, we compare how centralized and decentralized markets allocate risk. We begin with centralized trading to separate the changes relative to the competitive equilibrium due to noncompetitiveness itself versus market decentralization. The characterization of equilibrium for heterogeneous risk aversion—which, to the best of our knowledge, is new as is the uniqueness of the linear equilibrium

[11] Equilibrium schedules are optimal even if traders learn the independent value endowments q_i^0 (or equivalently, stochastic marginal utility intercepts, $\tilde{d} = d - \alpha\Sigma q_i^0$) of all other agents. The key to this ex post property of Bayesian Nash equilibrium is that permitting pointwise optimization—for each price—equilibrium demand schedules are optimal for any distribution of independent private information and are independent of agents' expectations about others' endowments. Equilibrium is *linear* if schedules have the functional form of $q_i(\cdot) = \alpha_0 + \alpha_{i,q} q_i^0 + \alpha_{i,p} p$. Strategies are not restricted to linear schedules; rather, it is optimal for a trader to submit a linear demand given that others do. The approach of analyzing the symmetric linear equilibrium is common in centralized market models (e.g., Kyle 1989; Vayanos 1999; Vives 2011). Our analysis does not assume equilibrium symmetry.

for many assets (divisible goods) in Proposition 1—is the key to the welfare effects of centralized and decentralized trading.

A. *Equilibrium and Allocation of Risk in Centralized Markets*

Suppose that all traders participate in a single exchange: $N(i) = N = \{(I, K)\}$ and $\mathbf{V}_{N(i)} = \Sigma = \mathbf{V}$ for all $i \in I$ (Example 1 (i)). In equilibrium, the (net) demand schedule of trader i equalizes his marginal utility with his marginal payment for each price,

$$d - \alpha_i \Sigma\big(q_i^0 + q_i\big) = p + \Lambda_i q_i, \tag{4}$$

where Λ_i measures the *price impact* of trader i in the exchanges in which he participates (i.e., "Kyle's lambda"). Here, Λ_i is the $K \times K$ Jacobian matrix of the inverse residual supply of trader i, which is defined by aggregation through the market clearing of the schedules submitted by other traders, $\{q_j(p)\colon \mathbb{R}^K \to \mathbb{R}^K\}_{j \neq i}$. The inverse of price impact is a common measure of *liquidity*: the lower the price impact, the smaller the price concession a trader needs to accept to trade, the more liquid the market. It follows from (4) that *if* trader i knew his price impact Λ_i, which is endogenous, he could determine his demand by equalizing his marginal utility and marginal payment pointwise. Let $q_i(\,\cdot\,, \Lambda_i)\colon \mathbb{R}^K \to \mathbb{R}^K$ be the schedule defined by pointwise optimization (4) for all prices p by trader i, given his assumed price impact Λ_i,

$$q_i(p, \Lambda_i) = (\alpha_i \Sigma + \Lambda_i)^{-1}(d - p - \alpha_i \Sigma q_i^0). \tag{5}$$

Equilibrium price impacts $\{\Lambda_i\}_i$ can now be determined by market clearing. Namely, the equilibrium condition requires that the price impact assumed by trader i in his pointwise optimization (5) is equal to the actual slope of his inverse residual supply, resulting from the aggregation of the other traders' submitted schedules. Proposition 1 shows that the system for equilibrium price impacts can be solved explicitly.

PROPOSITION 1 (Centralized Market Equilibrium): *A profile of demand schedules and price impacts* $\{q_i(\,\cdot\,, \Lambda_i), \Lambda_i\}_i$ *is an equilibrium in a centralized market if and only if*

(*i*) *each trader i submits schedule* (5), *given his price impact* Λ_i,

(*ii*) *trader i*'s *price impact is*

$$\Lambda_i = \Big(\textstyle\sum_{j \neq i} (\alpha_j \Sigma + \Lambda_j)^{-1}\Big)^{-1}, \quad i = 1, \ldots, I. \tag{6}$$

Furthermore,

(*iii*) *Equilibrium exists and is unique*;

(*iv*) *Trader i's price impact* Λ_i *is proportional to the covariance matrix* Σ;

$$(7)\quad \Lambda_i = \beta_i \alpha_i \Sigma \quad \textit{with} \quad \beta_i = \frac{2}{\alpha_i b - 2 + \sqrt{(\alpha_i b)^2 + 4}},$$

where $b \in \mathbb{R}_+$ *is the unique positive solution to* $\sum_j (\alpha_j b + 2 + \sqrt{(\alpha_j b)^2 + 4})^{-1} = 1/2$. Λ_i *is monotone decreasing in* α_i.

Conditions (i) and (ii) jointly provide an equivalent representation of Nash equilibrium in schedules: (i) traders optimize given their assumed price impacts, (ii) which are correct. Analyzing price impact directly will be useful for understanding the implications of noncompetitive behavior.[12] Using (5), trader i submits the schedule

$$(8)\qquad q_i(p) = \gamma_i (\alpha_i \Sigma)^{-1} (d - p - \alpha_i \Sigma q_i^0).$$

With positive price impact $\Lambda_i > 0$, trader i demands (or sells) less ($\gamma_i < 1$) than if he had submitted his competitive schedule; $\gamma_i \equiv \frac{1}{1+\beta_i}$ is the trader's noncompetitive *demand reduction* relative to his competitive demand ($\gamma_i = 1$) and β_i can be interpreted as price impact per unit of risk. That is, as $I \to \infty$, then $\Lambda_i \to 0$ for all i, and the competitive limit bid coincides with the inverse marginal utility, given the quasilinearity of the utility function. Panel A of Figure 1 illustrates. It depicts the best response net demand of trader i given his price impact $\Lambda_i > 0$ relative to his marginal utility. For prices such that $q_i > 0$, trader i is a buyer; for prices such that $q_i < 0$, he is a seller.

The central message of this paper is that, given the set of traders and assets, the centralized market may be inefficient in allocating risk relative to a decentralized market structure. Corollary 1 characterizes how centralized markets allocate risk and is the starting point for understanding why decentralized markets might be more efficient. Let $\gamma \equiv (\gamma_i)$ and $q^0 \equiv (q_i^0)$ be the vectors composed of γ_i and q_i^0, respectively.

COROLLARY 1 (Centralized Market Allocations and Prices): *Let* $\bar{\alpha} \equiv \frac{1}{I} \frac{\sum_j \gamma_j}{\sum_j \frac{\gamma_j}{\alpha_j}}$ *and*

$$(9)\qquad q^* \equiv \left(\sum_j \frac{\gamma_j}{\alpha_j}\right)^{-1} \sum_j \gamma_j q_j^0 = \bar{\alpha} \sum_j q_j^0 + \left(\sum_j \frac{\gamma_j}{\alpha_j}\right)^{-1} I \operatorname{cov}(\gamma, q^0).$$

Equilibrium trade and allocation of agent i are, respectively,

$$(10)\qquad q_i = \gamma^i (\alpha_i^{-1} q^* - q_i^0) \quad \textit{and} \quad q_i + q_i^0 = \gamma_i \alpha_i^{-1} q^* + (1 - \gamma_i) q_i^0,$$

and the vector of market-clearing prices is given by $p = d - \Sigma q^*$.

[12] For symmetric risk aversion, equilibrium from Proposition 1 coincides with that in Rostek and Weretka (2015a), which coincides with Kyle (1989, without nonstrategic traders and assuming independent values), Vayanos (1999), and Vives (2011). For a nonstrategic characterization of equilibrium in a general equilibrium setting (i.e., in terms of price and quantity levels, rather than demand functions), see Weretka (2011). Rostek and Weretka (2015a) introduce the Nash equilibrium representation and the equivalence result in Proposition 1 for centralized market games by formulating conditions on demand schedules and price impacts.

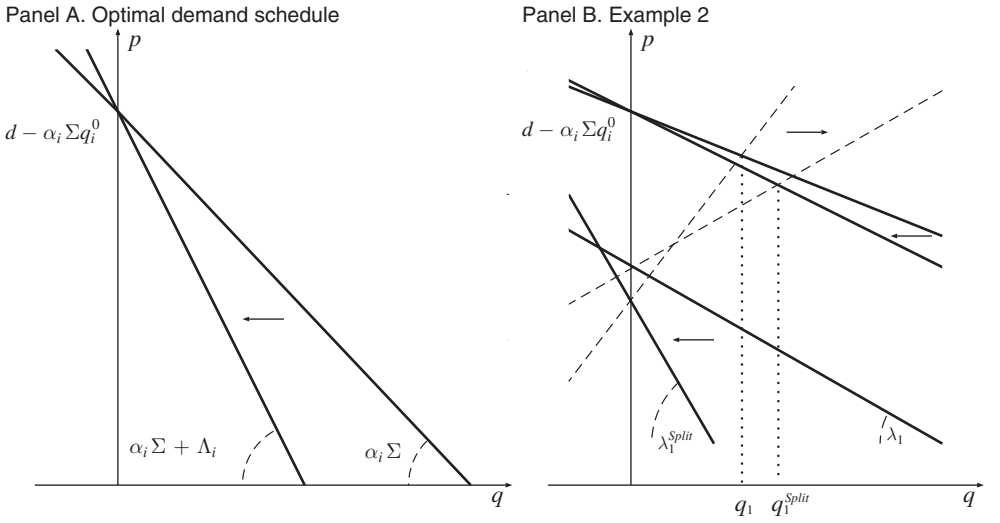

FIGURE 1

In the competitive centralized market, equilibrium allocation coincides with the *efficient allocation*, which maximizes total welfare (3) over the set of all feasible allocations $\{\{q_i\}_i : \sum_i q_i = \sum_i q_i^0\}$, or equivalently, minimizes the total utility loss due to risk exposure, and is given by

$$q_i^{**} = \alpha_i^{-1} q^{**}, \quad \text{where } q^{**} = \left(\sum_j \alpha_j^{-1}\right)^{-1} \sum_j q_j^0. \tag{11}$$

When traders have price impact, each trader's allocation is a combination of endowment risk q_i^0 and portfolio q^*, which is common to all traders. The equilibrium allocation differs from the competitive one in two ways. First, a trader retains a fraction γ_i of his initial endowment q_i^0. While γ_i can be characterized in terms of primitives, it is useful to relate it to price impact (using $\gamma_i \equiv (1/(1+\beta_i))$ and equation (7)): the smaller a trader's price impact, the closer γ_i is to 1, the less of his initial endowment the trader retains in equilibrium. Second, the common portfolio q^* that gets allocated to all agents differs from the efficient portfolio, unless traders risk preferences $\{\alpha_i\}_i$ are symmetric (then, $\gamma_i = \frac{I-2}{I-1}$ for all i). Equilibrium noncompetitiveness gives rise to a discrepancy between *systematic risk* Σq^{**}, which is defined by (11), independent of the market structure and represents risk nondiversifiable in the market, and the risk that is nondiversifiable in equilibrium, which we refer to as *aggregate risk*.

DEFINITION 1 (Aggregate Risk): *Aggregate risk $d - p$ represents the risk that is not diversified in equilibrium and corresponds to the risk premium in prices relative to the mean return d. In the centralized market, q^* is the aggregate risk portfolio: $d - p = \Sigma q^*$.*

In competitive markets, aggregate and systematic risk coincide and depend on the aggregate endowment $\sum_j q_j^0$ alone. In noncompetitive markets, aggregate risk is a function of equilibrium price impact. This is intuitive: what constitutes risk

that is nondiversifiable in equilibrium depends on price impact. A smaller γ_i lowers the contribution of agent i's endowment to aggregate risk and equilibrium prices. In equation (9), $\bar{\alpha}$ is the counterpart of aggregate risk aversion in the competitive market.

Example 2 shows that breaking up the centralized market into disjoint exchanges can increase welfare. It builds directly on and illustrates Corollary 1.

Example 2 (Splitting the Market Can Increase Welfare): Consider a market for one asset with variance $\sigma^2 = 1$ and three classes of agents of equal size $M \geq 3$, with risk aversion α_i in each class $i = 1, 2, 3$. If all agents' endowments were efficient, they would not trade (this holds for any market structure, Corollary 2), and adding or subtracting the efficient allocation would not change the conclusions. Therefore, without loss of generality, we may conveniently assume that the *aggregate endowment* $\sum_i q_i^0$ is 0, and hence, by (11), $q_i^{**} = 0$ for all i.

Suppose that each agent of classes 1 and 2 is endowed with $-q$ and q units of the asset, respectively, while class 3 has an endowment of 0, and $\alpha_1 < \alpha_2 \leq \alpha_3$. It is easy to see that the efficient allocation would have each agent of class 2 sell q units to class 1, and agents of class 3 would not trade, since they are already holding their efficient allocation.

However, in the noncompetitive centralized market, class 3 agents do trade in equilibrium, despite their efficient initial allocation, because the aggregate portfolio differs from the efficient one: $\alpha_2 > \alpha_1$ implies $\gamma_2 > \gamma_1$; hence by (9), $q^* > 0$. Price is below its competitive level, which makes it optimal for agents of class 3 to hold $q_3^0 + q_3 = \gamma_3 \alpha_3^{-1} q^* > 0$. The other classes do not diversify fully: $q_i^0 + q_i = (1 - \gamma_i) q_i^0 + \gamma_i \alpha_i^{-1} q^*$, $i = 1, 2$.

Consider next a decentralized market created by breaking up the centralized market into two exchanges, one for classes 1 and 2 and the other for class 3. In the second exchange, since class 3 agents have identical endowments, they do not trade, which is efficient; $q_3^0 + q_3^{Split} = 0$. In the first exchange, price impact increases strictly for all traders; we will show that an increase in price impact is a general result of decentralization (Theorem 2). However, having excluded class 3 agents, the relative price impacts (and weights $\{\gamma_i\}_i$) change so that the aggregate risk increases: denoting the aggregate risk portfolio in the first exchange by $q^{*,Split,1}$, we have $q^{*,Split,1} \equiv \left(\sum_{j \in 1} \frac{\gamma_j}{\alpha_j}\right)^{-1} \sum_{j \in 1} \gamma_j q_j^0 > q^*$. With sufficient heterogeneity in initial endowments (for any risk aversion), the increase is such that by buying a smaller fraction (due to the larger price impact) of the larger aggregate portfolio class 1 traders attain an allocation closer to the efficient allocation of 0 than in the centralized market; in equilibrium, $0 > q_1^0 + q_1^{Split,1} > q_1^0 + q_1$. Moreover, compared to the centralized market, the inefficiency is concentrated among class 1 and 2 traders. Given the concavity of utility, with sufficient heterogeneity in endowments, the total welfare increases (online Appendix D provides details).

Panel B of Figure 1 illustrates the corresponding shifts in demand and residual supply of class 1. In the centralized market, the residual supply of a class 1 agent aggregates the net demands of all other agents; in the split market, it aggregates the net demands of all other traders excluding class 3. In response to their larger price impact in the split market, $\lambda_1 < \lambda_1^{Split}$, class 1 agents reduce their net demands for all

prices. However, the residual supply of class 1 shifts having excluded class 3 agents from the exchange and class 1 agents buy more in the split market: $q_1^{Split} > q_1$.

Heterogeneity in risk aversions is crucial: with symmetric risk preferences, aggregate risk would be the same in the two market structures. While allocation would not be efficient due to price impact, we will show that when this is the only type of inefficiency, the centralized market is still the second best (Proposition 3). However, this need not be the case when some traders do not trade monotonically toward the efficient allocation, as this example illustrates: if some agents' initial endowments are close to efficient, in the centralized market they will trade inefficiently (class 3); conversely, if some agents' initial endowments are highly inefficient and their risk aversion is low (price impact is high), they will trade too little relative to what a decentralized market allows (class 1).

The result of Example 2 is striking. In centralized competitive markets, removing agents from the market cannot improve welfare for other agents; the competitive allocation is in the core. Equation (9) provides the first key observation: unless agents' risk preferences $\{\alpha_i\}_i$ are symmetric or agents' price impacts are 0 ($\gamma_i = 1$ for all i), *the aggregate risk portfolio q^* differs from the efficient portfolio q^{**}, which is independent of the market structure*. In particular, aggregate risk depends on the joint distribution of endowment risk $\{q_i^0\}_i$ and risk preferences (equilibrium price impact $\{\gamma_i\}_i$) rather than on the aggregate endowment alone.

The second main observation is that the centralized market allocates risk in a particular way. Namely, *in the centralized market*, *less risk-averse agents have greater price impact*: if $\alpha_1 < \cdots < \alpha_I$, then $\Lambda_1 > \cdots > \Lambda_I$; less risk-averse agents face a more risk-averse residual market, and therefore a less elastic residual supply (cf. equation (6)). Since less risk-averse agents will be reluctant to buy (or sell), equilibrium prices will be low (or high), and if their endowment happens to significantly differ from their efficient allocation, other agents will retain a large fraction of their nondiversified risk. Thus, *in noncompetitive markets*, *the efficiency of the centralized market allocation depends on the joint distribution of initial endowments and risk preferences*. This also suggests that a decentralized market can be more efficient than the centralized market if it reallocates risk so that more risk-averse agents attain allocation closer to the efficient one. We will next characterize how in a decentralized market, equilibrium price impact and the aggregate portfolio depend on the market structure.

B. *Equilibrium and Allocation of Risk in Decentralized Markets*

Theorem 1 shows that conditions analogous to those in Proposition 1 characterize equilibrium in the general model of decentralized markets. Consider the optimization problem (2) of a trader i who submits a demand schedule in exchanges $N(i)$, in which he participates,

$$q_i(p_{N(i)}, \Lambda_i) = (\alpha_i \mathbf{V}_{N(i)} + \Lambda_i)^{-1}(d_{N(i)} - p_{N(i)} - \alpha_i \mathbf{V}_{N(i)} q_i^0). \tag{12}$$

Recall that $\mathbf{V}$ is the covariance matrix of all assets traded on all exchanges and $\mathbf{V}_{N(i)}$ is its submatrix with rows and columns from exchanges $N(i) \subseteq N$. In analogy

to (5), trader i's demand equalizes his marginal utility and marginal payment, which depends on his price impact Λ_i in the exchanges in which he participates: the $N(i) \times N(i)$ Jacobian matrix $\Lambda_i = \left(\frac{\partial(p_1,\ldots,p_{N(i)})}{\partial(q_{i,1},\ldots,q_{i,N(i)})}\right) = \left(\frac{\partial p_\ell}{\partial q_{i,k}}\right)_{k,\ell}$. Entry (k,ℓ) represents the price change of asset ℓ that results from a marginal increase in the demanded quantity of asset k.

In equilibrium, trader i's price impact Λ_i must equal the slope of his residual inverse supply, which is defined by aggregation through market clearing of the schedules submitted by other traders, $\{q_j(p_{N(j)}) : \mathbb{R}^{N(j)} \to \mathbb{R}^{N(j)}\}_{j\neq i}$. In a decentralized market, traders' price impacts are of different dimensionality, correspond to different assets, and in general are not independent across exchanges, so the market-clearing condition cannot be written exchange-by-exchange. To apply market clearing to all assets in all exchanges, we use the procedure of *lifting*, which restores common dimensionality. For a given subset $N(i) \subseteq N$, decompose $\mathbb{R}^N = \mathbb{R}^{N(i)} \oplus \mathbb{R}^{N\setminus N(i)}$ as a direct sum of two subspaces corresponding to coordinates that agent i trades and those that he does not trade, where $\mathbb{R}^N$ is the space of asset holdings of dimension $\sum_{n\in N} K(n)$. Any symmetric matrix A can be decomposed into a block form

$$A = \begin{pmatrix} A_{i,i} & A_{i,-i} \\ A_{i,-i}^T & A_{-i,-i} \end{pmatrix}, \tag{13}$$

where $A_{i,i} = A_{N(i)}$ acts on subspace $\mathbb{R}^{N(i)}$, $A_{-i,-i} = A_{N\setminus N(i)}$ acts on the complementary subspace $\mathbb{R}^{N\setminus N(i)}$, and $A_{i,-i}$ is a rectangular block.

DEFINITION 2 (Lifting): *For any matrix $A_{i,i} \in \mathbb{R}^{N(i)\times N(i)}$, let $\bar{A}_{i,i}$ denote the lifted matrix which acts on $\mathbb{R}^N$, and with a slight abuse of notation, let $\bar{A}_{i,i}^{-1}$ denote its inverse:*[13]

$$\bar{A}_{i,i} \equiv \begin{pmatrix} A_{i,i} & 0 \\ 0 & 0 \end{pmatrix}, \quad \bar{A}_{i,i}^{-1} \equiv \begin{pmatrix} A_{i,i}^{-1} & 0 \\ 0 & 0 \end{pmatrix}. \tag{14}$$

In what follows, for simplicity of notation, we use $\bar{q}_i \in \mathbb{R}^N$ and $\bar{q}_i^0 \in \mathbb{R}^N$ to also denote the vectors $q_i \in \mathbb{R}^{N(i)}$ and $q_i^0 \in \mathbb{R}^{N(i)}$ "completed" by zeros in their $\mathbb{R}^{N\setminus N(i)}$ coordinates. In this notation, we can write the lifted demand schedule of agent i as $\bar{q}_i(p_{N(i)}) = (\alpha_j \bar{\mathbf{V}}_{N(j)} + \bar{\Lambda}_j)^{-1}(d - p - \alpha_i \mathbf{V}\bar{q}_i^0)$.

Treating assets traded in different exchanges as distinct assets and dealing with aggregation through lifting allows us to characterize equilibria in any decentralized market by two conditions: (i) each trader submits a schedule that equalizes his marginal utility and marginal payment given his price impact (i.e., submits $q_i(\cdot, \Lambda_i)$), and (ii) the price impact Λ_i in $q_i(\cdot, \Lambda_i)$ is correct (i.e., it equals the slope of the residual supply resulting from the aggregation of other traders' schedules, projected on the assets relevant for trader i). Recall that for two symmetric matrices A, B, matrix A is larger than B in the positive semidefinite order if $A - B$ is positive semidefinite; we write $A \geq B$.

[13] $A_{i,i}^{-1}$ is the Moore-Penrose pseudoinverse (Penrose 1955) of $A_{i,i}$ if $A_{i,i}$ is not invertible.

THEOREM 1 (Decentralized Market Equilibrium): *A profile of demand schedules and price impacts* $\{q_i(\cdot, \Lambda_i), \Lambda_i\}_i$ *is an equilibrium in a decentralized market with trader participation* $\{N(i)\}_i$ *if and only if*

(*i*) *each trader i submits schedule* (12), *given his price impact*,

(*ii*) *i's price impact* (*projected on the exchanges* $N(i)$ *where he participates*, *after lifting*) *is*

$$\Lambda_i = \left(\left(\sum_{j \neq i} (\alpha_j \bar{\mathbf{V}}_{N(j)} + \bar{\Lambda}_j)^{-1}\right)^{-1}\right)_{N(i)}, \quad i \in I. \tag{15}$$

Furthermore,

(*iii*) *Equilibrium exists and is locally unique*, *generically in risk aversions and the covariance matrix*.[14]

(*iv*) *For any decentralized market*, *the set of equilibrium price impact tuples has unique maximal and minimal elements in the sense of the positive semidefinite order. Equilibrium price impacts are positive semidefinite. If the covariance matrix* $\mathbf{V}_{N(i)}$ *is invertible for any* i, *price impacts are positive definite*.[15]

Although covariances $\mathbf{V}_{N(j)}$ and price impacts Λ_j are for "local" exchanges, when lifted, the same aggregation condition (6) as in the centralized market applies. Theorem 1 thus allows a direct comparison of equilibrium in decentralized and centralized markets. In contrast to centralized markets, *price impacts are generally not proportional to the fundamental covariance matrix*. In fact, Λ_i is not proportional to $\mathbf{V}_{N(i)}$ for generic $\mathbf{V}_{N(i)}$. The positive semidefiniteness of price impact is implied by equilibrium. By (15), if price impact were not positive semidefinite, $\alpha_j \mathbf{V}_{N(j)} + \Lambda_j$ would not be either for some j.[16] Then, for some portfolio y, $y^T(\alpha_j \mathbf{V}_{N(j)} + \Lambda_j)y < 0$ and buying an infinite amount of y would be optimal.

The double auction game—centralized and decentralized—can be equivalently seen as a game in which agents choose their demand slope $S_i \equiv (\alpha_i \mathbf{V}_{N(i)} + \Lambda_i)^{-1}$. Since these demand slopes are positive semi-definite, we can study them using the positive semi-definite order extended to tuples and say that one tuple is larger than another one if it is larger coordinate by coordinate. In the centralized market, the set of slope tuples $\{S_i\}_i$ is a lattice; this follows from the proportionality of price impact in the covariance so that $(\alpha_i \Sigma + \Lambda_i)^{-1} = \gamma_i(\alpha_i \Sigma)^{-1}$ and hence the order coincides with the natural order on $\mathbb{R}^I$. In decentralized markets, since the set of symmetric matrices is not a lattice with respect to the positive semidefinite order,[17]

[14] That is, for almost every positive definite matrix with respect to the measure induced by the Lebesgue measure on the set of positive definite matrices.

[15] Induced by the covariance matrix Σ and trader participation $\{N(i)\}_i$, $\mathbf{V}_{N(i)}$ can be singular only if agent i can trade the same asset in different exchanges. In this case, agents are indifferent about which exchange to trade in and so (replicas of) assets trade at the same prices.

[16] Indeed, if $\alpha_j \mathbf{V}_{N(j)} + \Lambda_j$ is positive semidefinite for each j, then so is $(\alpha_j \mathbf{V}_{N(j)} + \Lambda_j)^{-1}$ (because inversion preserves positive semi-definiteness), and hence so is also the right-hand side of (15).

[17] It is generally not possible to define the greatest lower bound and the least upper bound for a bounded set of positive semidefinite matrices. For example, consider $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Note that both $A \not\geq B$ and

Tarski's fixed point theorem cannot be applied to prove the existence of equilibrium and comparative statics. Nevertheless, we show that the price impact of each agent is monotone increasing (in this order) in the price impacts of others; standard arguments then give existence. Exploiting properties of positive definite matrices and refining the iterative procedure allows us then to prove existence of unique maximal and minimal elements in the set of equilibria.[18] We show that for any equilibrium in a less decentralized market, there is an equilibrium in the more decentralized market that has a larger price impact. All statements about more decentralized markets in Theorem 2 hold for any such equilibrium in these markets. In the sequel, we refer to either the minimal or maximal equilibria, which are unique.[19]

Let us remark that if a trader who knows his own utility knows his own price impact in the exchanges in which he participates, then by Theorem 1, his strategy $q_i(\cdot, \Lambda_i)$ would not be altered by knowledge of the market structure $\{N(i)\}_i$, the terms of trade in exchanges $N \backslash N(i)$, or even the submitted schedules or preferences of the traders in $N \backslash N(i)$. Despite the potential complexity of the trading environment, through the fixed-point condition (15), price impact Λ_i is the sufficient statistic for the optimality of trader i's schedule in exchanges $N(i)$, given the schedules of all traders $j \neq i$ in *all* exchanges $n \in N$ (i.e., $\Lambda_i \in \mathbb{R}^{N(i)\times N(i)}$ is sufficient for $\{q_j(\cdot): \mathbb{R}^{N(j)} \to \mathbb{R}^{N(j)}\}_{j\neq i}$). Thus, the opacity of decentralized markets is without loss of generality for equilibrium in trading environments with independent private values.

By Theorem 1, analysis of the equilibrium properties effectively reduces to studying the properties of the solution to the fixed point system (15) for price impacts. We illustrate the properties of this system with the following example.

Example 3 (Equilibrium Price Impacts): Consider a market with two exchanges for one asset and three classes of agents, with M_i agents in each class $i = 1, 2, 3$. Agents of class 1 trade only in exchange 1, agents of class 3 trade only in exchange 2, and the agent of class 2 ($M_2 = 1$) trades in both exchanges. Price impacts Λ_1 and Λ_3 are scalars, while Λ_2 is a matrix. The lifted price impacts and the covariance matrix for the market are given by

$$\bar{\Lambda}_1 = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \bar{\Lambda}_3 = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_3 \end{pmatrix}, \quad \bar{\Lambda}_2 = \Lambda_2 = \begin{pmatrix} \Lambda_2^{11} & \Lambda_2^{12} \\ \Lambda_2^{21} & \Lambda_2^{22} \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{pmatrix},$$

$A \not\leq B$ hold, because the positive semidefinite order is incomplete. By definition, matrix $C = \left(\begin{smallmatrix} a & c \\ c & b \end{smallmatrix}\right)$ is the least upper bound of A and B if $C \geq A$, $C \geq B$, and any other matrix C' satisfying $C' \geq A$, $C' \geq B$ also satisfies $C' \geq C$. However, $C \geq A$ and $C \geq B$ is equivalent to $a > 1$, $b > 2$, $(a-1)(b-2) \geq \max\{c^2, (c-1)^2\}$. Clearly, one can decrease a and increase b without violating these inequalities, which implies that C cannot be the least upper bound. Separately from the absence of the lattice order on strategies, an agent's utility differences are, in general, not monotone in the changes of strategies by others—the decentralized-market game is not supermodular.

[18] While Theorem 1 only establishes the existence of an equilibrium, we conjecture that equilibrium is globally unique. This is true in all of the examples in the paper and is confirmed by extensive numerical simulations. Equilibrium, defined as $\{q_i(\cdot, \Lambda_i), \Lambda_i\}_i$, is locally unique for generic parameters (this is a consequence of real analyticity of the equilibrium system; see, for example, Hugonnier, Malamud, and Trubowitz 2012).

[19] Let us relate our uniqueness results in Proposition 1 and Theorem 1 to that in Lambert, Ostrovsky, and Panov (2016), who consider a (centralized) game in which strategies are quantities (market orders). We analyze games in demand and supply functions. In that paper, there is a single asset and a single liquidity provider; thus, there is a scalar price impact, the same for all agents, and this one number solves a quadratic equation that has a unique positive solution, hence the unique equilibrium. By contrast, we have multiple assets, multiple exchanges, and multiple heterogeneous price impacts that are matrices and solve a system of nonlinear equations.

where σ^2 is the asset variance, and the fixed point system is

$$\begin{aligned}(16)\qquad \Lambda_1 &= \Big(\big((M_1-1)(\bar{\Lambda}_1+\alpha_1\bar{\mathbf{V}}_{11})^{-1}+(\Lambda_2+\alpha_2\mathbf{V})^{-1}\\ &\qquad +M_3(\bar{\Lambda}_3+\alpha_3\bar{\mathbf{V}}_{22})^{-1}\big)^{-1}\Big)_{11},\\ \Lambda_3 &= \Big(\big(M_1(\bar{\Lambda}_1+\alpha_1\bar{\mathbf{V}}_{11})^{-1}+(\Lambda_2+\alpha_2\mathbf{V})^{-1}\\ &\qquad +(M_3-1)(\bar{\Lambda}_3+\alpha_3\bar{\mathbf{V}}_{22})^{-1}\big)^{-1}\Big)_{22},\\ \Lambda_2 &= \big(M_1(\bar{\Lambda}_1+\alpha_1\bar{\mathbf{V}}_{11})^{-1}+M_3(\bar{\Lambda}_3+\alpha_3\bar{\mathbf{V}}_{22})^{-1}\big)^{-1}.\end{aligned}$$

We can rewrite the equation for Λ_2 as follows: by the definition of lifting,

$$(17)\qquad \Lambda_2 = \begin{pmatrix} M_1^{-1}(\Lambda_1+\alpha_1\mathbf{V}_{11}) & 0\\ 0 & M_3^{-1}(\Lambda_3+\alpha_3\mathbf{V}_{11})\end{pmatrix}.$$

We will show (Theorem 2) that equilibrium price impacts of traders in both exchanges decrease when the number of traders increases in exchange 1 or the connected exchange 2, or their risk aversion decreases: the less risk-averse the traders of class 1 are and the larger their number M_1, the more liquidity they provide to class 2, lowering Λ_2, and in turn, the more liquidity the trader of class 2 provides to class 3, lowering Λ_3.

The interdependence among price impacts is anticipated by Theorem 1. In general, the equilibrium price impact of trader i in exchanges $N(i)$ depends, and positively so, not only on the price impacts of other traders in exchanges $N(i)$ but also on the price impacts of the traders in all other exchanges $N\backslash N(i)$ (in the same connected component), including those with whom he is linked only through counterparties (e.g., classes 1 and 3 in Example 3). More generally, a trader's price impact in a decentralized market also depends on the price impacts of the traders in exchanges in which his counterparties do not participate.

Example 3 also shows that even if the total number of traders is large, when trading is decentralized, traders generally have nonnegligible price impact in the exchanges in which they participate. In general, when agents have access to a perfectly liquid exchange n, they will have strictly positive price impact in the other exchanges $N(i)\backslash\{n\}$ in which they participate.

Corollary 2 characterizes equilibrium trades and prices in decentralized markets. Corollary 4 in online Appendix D explicitly characterizes allocations for markets with one asset.

COROLLARY 2 (Decentralized Market Allocations and Prices): *Let*

$$(18)\qquad \mathbf{Q}^* = \Big(\sum_j(\alpha_j\bar{\mathbf{V}}_{N(j)}+\bar{\Lambda}_j)^{-1}\Big)^{-1}\sum_j\bar{\Gamma}_j q_j^0,$$

and $\Gamma_i \equiv (\alpha_i \mathbf{V}_{N(i)} + \Lambda_i)^{-1} \alpha_i \mathbf{V}_{N(i)}$. *The equilibrium allocation of agent i is*

$$q_i + q_i^0 = (\alpha_i \mathbf{V}_{N(i)} + \Lambda_i)^{-1} \mathbf{Q}^*_{N(i)} + (\mathrm{Id} - \Gamma_i) q_i^0, \tag{19}$$

and the vector of market-clearing prices is given by $p = d - \mathbf{Q}^*$. *If the initial allocation is efficient,* $q_i^0 = q_i^{**}$ *for all i, then there is no trade.*

As in the noncompetitive centralized market equilibrium, each trader is exposed to endowment and aggregate risks. The term $\mathbf{Q}^*$ is the decentralized market counterpart of the aggregate risk Σq^*; it represents the risk premium $d - p = \mathbf{Q}^*$ in prices and is defined by market clearing applied to all traders' demands.[20,21] Like Σq^*, $\mathbf{Q}^*$ is also in risk units; unlike Σq^*, $\mathbf{Q}^*$ is not separable in the covariance. In addition to the effects brought by noncompetitiveness itself ($\Lambda_i > 0$)—i.e., allocation retains endowment risk, aggregate and systematic risk differ—decentralized trading makes both the risk nondiversifiable in equilibrium $\mathbf{Q}^*$ and a trader's equilibrium demand reduction $\mathrm{Id} - \Gamma_i$ depend on the market structure. Namely, both are a function of asset covariance, and $\mathbf{Q}^*$ also depends on distribution of endowments across exchanges. As a result, the quantity of aggregate risk allocated to a trader in the exchanges in which he participates generally depends on the preferences of agents and the covariance of assets from all exchanges. We should note that the dependence of risk $\mathbf{Q}^*_{N(i)}$ and weight $\mathrm{Id} - \Gamma_i$ on the covariances in exchanges $N \backslash N(i)$ as well as $N(i)$ arises in a decentralized market because price impact Λ_i is no longer proportional to covariance $\mathbf{V}_{N(i)}$. Example 4 illustrates how the properties of aggregate risk differ in decentralized and centralized markets.

Example 4 (Aggregate Risk): Consider the market from Example 2, and suppose that one of the agents of class 1, denoted by $1d$ (a dealer), can trade in both exchanges 1 and 2. The market is then less decentralized than the split market (cf. Definition 3), and agent $1d$ serves as an intermediary between class 2, class 3, and other agents of class 1. Using Example 3, the price impact of trader $1d$ is diagonal,

$$\Lambda_{1d} = \begin{pmatrix} (M-1)(\Lambda_1 + \alpha_1)^{-1} + M(\Lambda_2 + \alpha_2)^{-1} & 0 \\ 0 & M(\Lambda_3 + \alpha_3)^{-1} \end{pmatrix}^{-1}.$$

[20] The linearity of equilibrium prices in endowments leads us to interpret Corollary 2 in CAPM terms. A decentralized market CAPM holds, even with strategic traders whose price impact is nonnegligible. In each exchange, expectations of asset payoffs lie on a security-market line defined by an agent-specific portfolio. With many assets, agents diversify risk through different (and multiple) funds, which depend on their participation in exchanges. See Malamud and Rostek (2016) for a decentralized market CAPM.

[21] Malamud and Rostek (2016) show that when exchanges are linked by one trader, prices of the same asset generally differ; with two or more common participating traders, price impacts, and prices for the same asset equalize between exchanges and are generically noncompetitive. In turn, the presence of one (or many) trader(s) who can engage in pure riskless arbitrage would not affect the main conclusions. In fact, this often is not possible, for instance, in dealer-intermediated markets. Even if an arbitrageur could place buying and selling orders in two exchanges, in a noncompetitive market, he would have price impact. A large enough round-trip order would change prices and result in strictly negative profits. That is, in contrast to competitive markets, profits from arbitrage are not infinite, but bounded. Thus, unlike the competitive model, sufficient fixed entry costs can discourage outside investors from arbitraging the liquidity effect. In practice, entry costs involve not only explicit trading costs but also costs associated with learning and monitoring the characteristics of particular stocks.

As in Example 2, we assume that class 1 agents are endowed with $-q$ units of the asset, class 2 agents are endowed with q units, and agents of class 3 have zero initial endowment. Then,[22]

$$\sum_j \bar{\Gamma}_j q_j^0 = \begin{pmatrix} (M\Gamma_2 - (M-1)\Gamma_1)q \\ 0 \end{pmatrix} + \Gamma_{1d}\begin{pmatrix} -q/2 \\ -q/2 \end{pmatrix},$$

and Corollary 2 implies that the aggregate risk is given by

$$\mathbf{Q}^* = \begin{pmatrix} \mathbf{Q}_1^* \\ \mathbf{Q}_2^* \end{pmatrix} = \left(\begin{pmatrix} (M-1)(\Lambda_1+\alpha_1)^{-1} + M(\Lambda_2+\alpha_2)^{-1} & 0 \\ 0 & M(\Lambda_3+\alpha_3)^{-1} \end{pmatrix} + (\Lambda_{1d} + \alpha_1 \mathbf{V})^{-1} \right)^{-1} \times \left(\begin{pmatrix} (M\Gamma_2 - (M-1)\Gamma_1)q \\ 0 \end{pmatrix} + \Gamma_{1d}\begin{pmatrix} -q/2 \\ -q/2 \end{pmatrix} \right), \quad \text{where}$$

$$\mathbf{V} = \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{pmatrix}.$$

Observe that (i) the aggregate risk is common to all traders in a given exchange $n = 1, 2$; however, (ii) $\mathbf{Q}_n^*$ is a weighted average of the endowments of *all* agents in the market (in a connected component of the graph on exchanges). In addition, (iii) since the exchange-specific coordinates $\mathbf{Q}_1^*, \mathbf{Q}_2^*$ of the aggregate risk vector generally differ, the same asset trades at different prices in different exchanges. In particular, (iv) in contrast to the centralized market (Definition (1)), the aggregate risk is not separable in the covariance $\mathbf{V}$ due to the nonproportionality of price impact to $\mathbf{V}$ in decentralized markets; contrary to the standard CAPM, the risk premium cannot be written as $\mathbf{V}\tilde{Q}$ for some vector $\tilde{Q}$—otherwise, prices of identical assets would be equal across exchanges. Indeed, if $\mathbf{V} = \mathbf{1}$, all coordinates of $\mathbf{V}\tilde{Q}$ are equal.[23]

The equilibrium allocation of risk in this intermediated market is given by

$$\begin{aligned} q_1^0 + q_1 &= (\Lambda_1 + \alpha_1)^{-1}\mathbf{Q}_1^* - (1-\Gamma_1)q, \\ q_2^0 + q_2 &= (\Lambda_2 + \alpha_2)^{-1}\mathbf{Q}_1^* + (1-\Gamma_2)q, \\ q_3^0 + q_3 &= (\Lambda_3 + \alpha_3)^{-1}\mathbf{Q}_2^*, \\ q_{1d}^0 + q_{1d} &= (1,1)\cdot\left((\Lambda_{1d} + \alpha_1\mathbf{V})^{-1}\begin{pmatrix} \mathbf{Q}_1^* \\ \mathbf{Q}_2^* \end{pmatrix} + (\mathrm{Id} - \Gamma_{1d})\begin{pmatrix} -q/2 \\ -q/2 \end{pmatrix} \right), \end{aligned} \tag{20}$$

[22] The split of endowment $-q$ of trader $1d$ across the two exchanges can be arbitrary, because endowment affects only the marginal utility for the first unit, which is the same in both exchanges (equation (12)).

[23] $\mathbf{Q}_{N(i)}^*$ and $\mathbf{Q}_n^*$ represent how aggregate risk $\mathbf{Q}^*$ affects a trader or an exchange, respectively. In the presence of shocks, these trader-specific and exchange-specific risks can be seen as measuring systemic risk, various formalizations of which in the literature share its dependence on equilibrium and position in the network.

where the multiplication by $(1,1)$ in the last line sums the holdings of agent $1d$ in exchanges 1 and 2. The intermediary trades both components $\mathbf{Q}_1^*$, $\mathbf{Q}_2^*$ of aggregate risk, whereas all other traders buy or sell its distinct components.

It follows from Corollary 2 that decentralized trading changes agents' ability to diversify in two ways (cf. Proposition 1): aggregate risk (prices) changes, since agents now acquire exchange-dependent components of $\mathbf{Q}^*$, which may make it optimal for some to trade closer to their efficient allocation; and equally risk-averse agents with identical endowments need not diversify in the same way. In Example 2, the component of aggregate risk for class 1 (the natural buyers) increases, thus lowering prices, which leads class 1 to buy more, while the natural sellers (classes 2 and 3) sell less. Additionally, classes 2 and 3 do not equalize their equilibrium allocation of aggregate risk $\gamma_i \alpha_i^{-1} q^{Split,n}$, $n = 1,2$, as they would in the centralized market.

III. Market Decentralization: Price Impact and Welfare

Before we analyze how market decentralization affects welfare, we first examine how it affects traders' price impacts. Let us make precise how market structures and price impacts are ranked.

DEFINITION 3 (More Decentralized Than): *Fix the set of traders and assets* (I,K). *Consider two markets with the sets of exchanges* $\mathcal{M} = \{(I(n),K(n))\}_n$ *and* $\mathcal{M}' = \{(I(n'),K(n'))\}_{n'}$. *We say that the market* $\mathcal{M}'$ *is* more decentralized than $\mathcal{M}$ *if for any exchange* n' *in* $\mathcal{M}'$, *there exists an exchange* n *in* $\mathcal{M}$ *such that* $I(n') \subseteq I(n)$ *and* $K(n') \subseteq K(n)$.

Thus, a more decentralized market restricts the participation of some agents in trading, with respect to traders or assets. By definition, any decentralized market is more decentralized than the centralized market. Suppose that $\mathcal{M}'$ is more decentralized than $\mathcal{M}$, and let $\{\Lambda_i^{\mathcal{M}'}\}_i$ and $\{\Lambda_i^{\mathcal{M}}\}_i$ be the corresponding equilibrium price impacts (which are symmetric and positive semidefinite, by Theorem 1). We say that the price impact tuple $\{\Lambda_i^{\mathcal{M}}\}_i$ is smaller than $\{\Lambda_i^{\mathcal{M}'}\}_i$ and write $\{\Lambda_i^{\mathcal{M}}\}_i \leq \{\Lambda_i^{\mathcal{M}'}\}_i$ if for all i,

$$\left(\Lambda_i^{\mathcal{M}}\right)_{N'(i)} \leq \Lambda_i^{\mathcal{M}'}.$$

If price impacts are higher in the positive semidefinite order,[24] the agents behave as if the assets they are trading are riskier. Recall that if two symmetric positive semidefinite matrices A, B satisfy $A \geq B$, then the diagonal elements satisfy $A_{ii} \geq B_{ii}$ for all i (however, no implication for the ordering of the off-diagonal elements follows), $A_{N(i)} \geq B_{N(i)}$ for any i, and $A^{-1} \leq B^{-1}$.

[24] One cannot generally conclude that the dealer's price impact in exchange 2 is lower than in the split market: one can only compare the price impact in the exchanges in which the agent participates in both market structures (cf. Definition 4 in the online Appendix). In particular, it could happen that $\Lambda_{1d,22}^{DM} > \lambda_{1d}^{Split}$ if the risk aversion of class 3 agents is sufficiently low.

A. *Price Impact in Decentralized Markets*

Given a map Λ from the set of positive semi-definite covariance matrices into the set of positive semi-definite matrices, we say that this map is concave if $\Lambda(0.5(\mathbf{V}+\mathbf{V}')) \geq 0.5(\Lambda(\mathbf{V})+\Lambda(\mathbf{V}'))$ for any positive semi-definite matrices $\mathbf{V}$, $\mathbf{V}'$. Similarly, a map Λ from $\mathbb{R}^I_+$ to the set of positive semi-definite matrices is concave if $\Lambda(0.5(\alpha+\alpha')) \geq 0.5(\Lambda(\alpha)+\Lambda(\alpha'))$ for any $\alpha, \alpha' \in \mathbb{R}^I_+$.

THEOREM 2 (Price Impact Monotonicity and Concavity): *The following is true:*

(*i*) *Fix the set of traders and assets* (I, K). *If* $\mathcal{M}'$ *is more decentralized than* $\mathcal{M}$, *then* $\{\Lambda_i^{\mathcal{M}'}\}_i \geq \{\Lambda_i^{\mathcal{M}}\}_i$.

(*ii*) *The equilibrium price impact tuple* $\{\Lambda_i\}_i$ *is increasing and concave in risk aversion* $\{\alpha_j\}_j$ *and the covariance matrix* Σ *and is decreasing in the number of agents in the market.*

(*iii*) *Fix the market structure* $\mathcal{M}$:

(*a*) *If* $N(i) = N(j)$ *and* $\alpha_i \leq \alpha_j$, *then* $\Lambda_i \geq \Lambda_j$.

(*b*) *If* $N(i) \supset N(j)$ *and trader j participates in a single exchange for one asset, then* $\alpha_i \leq \alpha_j$ *implies that the equilibrium price impact of trader i in exchange* $N(j)$ *is larger than that of trader j*,

$$(\Lambda_i)_{N(j)} \geq \Lambda_j .$$

Theorem 2 reports two complementarity results on how the interconnectedness *among exchanges* $\{N(i)\}_i$ affects a trader's equilibrium price impact in exchanges $N(i)$: changes in either the market structure or the characteristics of traders or assets that lower any trader's price impact in some exchange lower the price impacts of all traders in all directly and indirectly connected exchanges.

Part (i) of Theorem 2 considers the effect of changes in the market structure on equilibrium liquidity. Theorem 2 implies that creating a new exchange for a subset of agents which operates along with the existing exchanges always (weakly) lowers the price impacts (improves liquidity) in all exchanges. Indeed, if agents participate in a new exchange, the price impact of these agents in their existing exchanges decreases, which lowers the price impact of their counterparties in those exchanges, and so on. In turn, splitting an exchange (e.g., Example 2) always weakly increases price impact. Moreover, this holds for any asset structure in the new market. In general, price impact is monotone in both the set inclusion of traders and assets and is thus minimal in the centralized market. It follows from Theorem 2 that the lowest price impact that agents I who trade assets K can achieve occurs when all agents participate in all potential exchanges—a market structure equivalent to a centralized market.[25]

[25] Malamud and Rostek (2016) show that sufficient trader participation among exchanges suffices for liquidity to be as high as in the centralized market.

The liquidity effects of market decentralization (i.e., lowering trader participation) characterized by Theorem 2 hold for any number and risk aversion of traders and any covariance matrix of assets in the exchanges before and after the change.

Part (ii) of Theorem 2 deals with the impact of varying the primitive characteristics of traders and assets. Relative to the observation that equilibrium price impacts are interdependent across exchanges (Theorem 1), Theorem 2 demonstrates a general *complementarity* of price impacts in a decentralized market: a trader's price impact in exchanges $N(i)$ depends *positively* (in the sense of the positive semidefinite order) on the market characteristics in both exchanges $N(i)$ and $N\backslash N(i)$. Intuitively, an agent's price impact from increasing his trade in exchanges $N(i)$ represents the price concessions required for other agents $j \neq i$ in exchanges $N(i)$ to absorb the trade. With decreasing marginal utility ($\alpha_j > 0$), more risk-averse counterparties j in exchanges $N(i)$ demand larger price concessions to compensate for the trade's impact on their own marginal utilities (cf. the first-order condition (4)), thus making residual supply less elastic, and hence price impact larger, for all other agents in $N(i)$. In addition, when trading is decentralized, the fewer and more risk-averse agent i's counterparties' trading partners are in exchanges $N\backslash N(i)$, the larger the price impacts of those counterparties j in their exchanges $N(j)$ and the larger the price concessions they require in exchanges $N(i)$.

The *concavity* of price impact in the fundamental asset covariance contrasts sharply with centralized markets, where price impact is proportional to the covariance (Proposition 1). The concavity holds because the riskiness of assets in an exchange in which trader i participates affects his incentives through the *residual riskiness* of these assets, net of the risk that can be diversified in other exchanges by i and his counterparties. Mathematically, a Gaussian conditioning argument applies. Denoting by $S \equiv \sum_j (\alpha_j \bar{\mathbf{V}}_{N(j)} + \bar{\Lambda}_j)^{-1}$ the slope of the aggregate (net) market demand, the condition that price impact is the harmonic mean of (lifted) inverse demand slopes can be written as follows:

$$\Lambda_i = \left(S_{-i}^{-1}\right)_{N(i)} = \left(\left(\left(S^{-1}\right)_{N(i)}\right)^{-1} - \left(\alpha_i \mathbf{V}_{N(i)} + \Lambda_i\right)^{-1}\right)^{-1}, \tag{21}$$

where $\left(\left(S^{-1}\right)_{N(i)}\right)^{-1} = S_{i,i} - S_{i,-i} S_{-i,-i}^{-1} S_{i,-i}^{T}$, using the notation for S from (13).[26] That is, the projection that defines the residual risk of the assets traded in exchanges $N(i)$ is endogenous, as it depends on price impacts. Equation (21) is equivalent to

$$\left(\Lambda_i^{-1} + \left(\alpha_i \mathbf{V}_{N(i)} + \Lambda_i\right)^{-1}\right)^{-1} = \left(S^{-1}\right)_{N(i)},$$

and since the harmonic mean function $f(x) = \left(x^{-1} + (v + x)^{-1}\right)^{-1}$ is convex, its functional inverse is concave, hence Λ_i is (weakly) concave as a function of S^{-1}. In the centralized market, since S^{-1} is proportional to $\mathbf{V} = \Sigma$, so is Λ_i. In a decentralized market, since price impact depends on the residual risk in $\left(\left(S^{-1}\right)_{N(i)}\right)^{-1}$ rather than the fundamental risk Σ, proportionality is absent and the concavity matters.

[26] $S_{i,i} - S_{i,-i} S_{-i,-i}^{-1} S_{i,-i}^{T}$ is a shorted operator. The proof of Theorem 2 is based on the monotonicity and concavity properties of the matrix harmonic mean that are derived in Anderson (1971) and Anderson and Duffin (1969) using the theory of shorted operators.

Notably, the concavity implies that lower risk (Σ) or lower risk aversion $\{\alpha_i\}_i$ play a greater role in the determination of a trader's price impact than high-risk assets and highly risk-averse traders. Having access to such low-risk assets or low risk-averse traders—whether directly in one's own exchanges $N(i)$ or indirectly via interconnections with one's counterparties' exchanges $N\backslash N(i)$—lowers the price impact of i's trades in $N(i)$ more than access to high-risk assets or highly risk-averse traders would.

Part (iii) of Theorem 2 provides a link between price impact and the participation of different agents. It implies that one can make a systematic prediction about the link between the trader's "position" in the market and price impact: if more connected agents are weakly less risk-averse, they have higher price impact. This is intuitive: since the price impact of each agent in a given exchange is determined by the risk exposure of the other agents in that exchange, the ability of the more connected trader i to diversify risk in exchanges $N(i)\backslash N(j)$ lowers his risk exposure in exchanges $N(j)$ relative to trader j's exposure, and thus the price impact of j in exchanges $N(j)$.

B. *Welfare in Decentralized Markets*

By Theorem 2, the centralized market minimizes price impact for all traders. Nevertheless, as Example 2 shows, a decentralized market can give rise to strictly higher welfare with the same traders and assets. To examine more systematically the source of the welfare gains from decentralization, we begin with a characterization of indirect utility as a function of price impact, substituting equilibrium trade $q_i = (\alpha_i \mathbf{V}_{N(i)} + \Lambda_i)^{-1}(\mathbf{Q}^*_{N(i)} - \alpha_i \mathbf{V}_{N(i)} q_i^0)$ and price into agents' utility function $U_i(q_i) = d^T_{N(i)}(q_i^0 + q_i) - \frac{\alpha_i}{2}(q_i^0 + q_i)^T \mathbf{V}_{N(i)}(q_i^0 + q_i) - (p_{N(i)})^T q_i$:

$$U_i^{\mathcal{M}}(\Lambda_i; q_i^0) = \underbrace{(q_i^0)^T d^T_{N(i)} - \frac{1}{2}(q_i^0)^T \alpha_i \mathbf{V}_{N(i)} q_i^0}_{\textit{Utility without trade}} + \underbrace{q_i^T(\tfrac{1}{2}\alpha_i \mathbf{V}_{N(i)} + \Lambda_i) q_i}_{\textit{Equilibrium surplus from trade}}. \tag{22}$$

To interpret, in any market structure—centralized or decentralized, competitive or noncompetitive—equilibrium utility from trade derives from the risk premium benefit in the payment $q_i \cdot (d^T_{N(i)} - p) = q_i \cdot \mathbf{Q}^*_{N(i)}$ net of the utility cost of buying risky assets $-\alpha_i q_i^T \mathbf{V}_{N(i)} q_i$. In the competitive centralized market, $\mathbf{Q}^* = \Sigma q^{**}$ and the equilibrium utility surplus from trade is $q_i^T(\frac{1}{2}\alpha_i \Sigma) q_i = q_i^T(\alpha_i \Sigma - \frac{1}{2}\alpha_i \Sigma) q_i$. In a noncompetitive market, price impact exposes a trader to additional risk due to others' equilibrium behavior, hence the *marginal equilibrium utility per unit of quantity traded* is $(\alpha_i \mathbf{V}_{N(i)} + \Lambda_i) - \frac{1}{2}\alpha_i \mathbf{V}_{N(i)}$ in the surplus of equation (22).

Proposition 2 shows that the (non)proportionality of price impact in fundamental risk $\mathbf{V}_{N(i)}$ has important implications for welfare effects of market decentralization. It is useful to write equilibrium trade of trader i in utility (22) as $q_i = S_i(\mathbf{Q}^*_{N(i)} - \alpha_i \mathbf{V}_{N(i)} q_i^0)$, where $S_i \equiv (\alpha_i \mathbf{V}_{N(i)} + \Lambda_i)^{-1}$ is his demand slope and $(\mathbf{Q}^*_{N(i)} - \alpha_i \mathbf{V}_{N(i)} q_i^0)$ can be interpreted as his gains from trade (given that traders have price impact):

$$\begin{aligned} U_i^{\mathcal{M}}(\Lambda_i; q_i^0) = {} & (q_i^0)^T d^T_{N(i)} - \frac{1}{2}(q_i^0)^T \alpha_i \mathbf{V}_{N(i)} q_i^0 \\ & + \left(\mathbf{Q}^*_{N(i)} - \alpha_i \mathbf{V}_{N(i)} q_i^0\right)^T \Upsilon_i(\Lambda_i)\left(\mathbf{Q}^*_{N(i)} - \alpha_i \mathbf{V}_{N(i)} q_i^0\right), \end{aligned} \tag{23}$$

where matrix $\Upsilon_i(\Lambda_i) \equiv S_i - \frac{1}{2}S_i\alpha_i\mathbf{V}_{N(i)}S_i$ is *the marginal utility per unit of risk*, since both quantities in $(\mathbf{Q}^*_{N(i)} - \alpha_i\mathbf{V}_{N(i)}q_i^0)$ are expressed in units of fundamental risk $\alpha_i\mathbf{V}_{N(i)}$.

PROPOSITION 2 (Marginal Equilibrium Utility): *Suppose that market $\mathcal{M}'$ is more decentralized than $\mathcal{M}$, and let $\Lambda_i^{\mathcal{M}}$ and $\Lambda_i^{\mathcal{M}'}$ be the corresponding price impacts. Then*,

(*i*) *If $\Lambda_i^{\mathcal{M}}$ is proportional to $\mathbf{V}_{N(i)}$, then $\bar{\Upsilon}_i\left(\Lambda_i^{\mathcal{M}'}\right) \leq \bar{\Upsilon}_i\left(\Lambda_i^{\mathcal{M}}\right)$;*

(*ii*) *In general, $\bar{\Upsilon}_i\left(\Lambda_i^{\mathcal{M}'}\right) \not\leq \bar{\Upsilon}_i\left(\Lambda_i^{\mathcal{M}}\right)$, i.e., $\bar{\Upsilon}_i\left(\Lambda_i^{\mathcal{M}}\right)$ need not decrease as the market becomes more decentralized.*

Proposition 2 presents the paper's two main implications for welfare. Using (23), by part (i), a *necessary* condition for a decentralized market to increase utility relative to the centralized market is that the components of the aggregate risk $\mathbf{Q}^*$ in $(\mathbf{Q}^*_{N(i)} - \alpha_i\mathbf{V}_{N(i)}q_i^0)$ must change in some exchanges so that the equilibrium allocation can become closer to the efficient one. Indeed, if aggregate risk is the same as in the centralized market, equation (23) and part (i) of Proposition 2 imply that every agent is better off in the centralized market. Example 2 illustrates part (i), in which price impact in the less decentralized market is proportional to the covariance, as is the case in the centralized market: the element of $\mathbf{Q}^*$ in exchange 1 increases, and class 1 allocation increases and is more efficient; the element of $\mathbf{Q}^*$ in exchange 2 decreases, and class 3 allocation decreases, which is efficient. Section IIIC further examines this necessary condition. In turn, using (23), part (ii) of Proposition 2 indicates that making a decentralized market more decentralized might improve welfare in the *Pareto* sense, *even if it does not affect aggregate risk*. We will show in Example 5 that this is the case and examine the new effect in Section IIID.

Part (i) of Proposition 2 implies that for market decentralization to increase welfare, *risk sharing among traders* must improve. Part (ii) implies that decentralizing a market can also increase welfare by improving *diversification*, since in a decentralized market, the relative weights across assets in a trader's equilibrium portfolio may differ from those in the efficient portfolio. Indeed, a risk-efficient portfolio is given by $\alpha_i\mathbf{V}_{N(i)}^{-1}(d-p)$, while the agent buys $(\Lambda_i + \alpha_i\mathbf{V}_{N(i)})^{-1}(d-p)$. If $\Lambda_i = \alpha_i\beta_i\mathbf{V}_{N(i)}$ is proportional to the fundamental covariance matrix, diversification weights are efficient while risk allocation is imperfect, as the agent buys only a fraction $1/(1+\beta_i)$ of the efficient portfolio. However, when Λ_i is not proportional to $\mathbf{V}_{N(i)}$, both risk allocation and diversification are imperfect. While Theorem 2 guarantees that making the market more decentralized always reduces the number of units of the risky assets the agents are willing to absorb for every price $(\Lambda_i^{\mathcal{M}'} \geq \Lambda_i^{\mathcal{M}})$, Proposition 2(ii) implies that diversification of an individual trader can actually become more efficient in a more decentralized market, and the latter effect can potentially offset the former in markets with multiple assets. Mathematically, without proportionality, a change in the market structure affects how trades of correlated assets substitute in creating utility, i.e., $(\Upsilon_i)_{n,m}$, $n \neq m$ (Example 5 gives the proof of part (ii)).

C. *When Does a Decentralized Market Increase Welfare?*

At the primitive level, heterogeneity in risk preferences is central to welfare gains from market decentralization. Proposition 3 shows that when all traders care equally about risk, the centralized market maximizes total welfare *for any endowment distribution*, regardless of the asset structure in a decentralized market.

PROPOSITION 3 (Welfare and Heterogeneity in Risk Aversion): *If traders' risk aversions* $\{\alpha_i\}_i$ *are sufficiently close, then the total welfare in the centralized market is strictly higher than in any decentralized market with the same traders and assets for any endowments: there exists* $\varepsilon > 0$ *such that total welfare in the centralized market is strictly higher than in a decentralized market for any endowments whenever* $(\max_i \alpha_i / \min_i \alpha_i) < 1 + \varepsilon$.

Intuitively, recall that a decentralized market trade-offs larger price impact with the potential benefit from reallocating risk among traders with different risk preference by letting agents trade distinct components of aggregate risk. Nevertheless, when traders' risk preferences are close, the benefit from reallocating risk is small. Differential allocation merely increases dispersion of allocations, thus reducing welfare.

Proposition 4 shows that when some traders' risk preferences are sufficiently heterogeneous, the welfare-maximizing market structure is decentralized for some endowment distributions.[27]

PROPOSITION 4 (Splitting the Market Can Increase Welfare): *Let* $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_I$. *Suppose that* $I \geq 7$ *and* α_{I-3}/α_3 *is sufficiently large. Then, there exists an open set of initial endowments and a partition of the set of traders into two exchanges such that the total welfare in the split market is higher than that in the centralized market.*

By Proposition 2, for welfare to increase with decentralized trading, the joint profile of risk preferences and endowment risk must be such that the increase in some agents' $\mathbf{Q}^*_{N(i)}$ is sufficiently strong to countervail the higher price impact (lower $\{\Upsilon_i\}_i$). Welfare gains from decentralization exist for some endowment distributions even if the number of traders is large—the condition on the heterogeneity in risk aversion in Proposition 4 needs to hold only for a few lowest- and highest-risk-preference traders. For the same reason, for a market decentralization to increase welfare, the endowments need not become extreme as the number of traders grows. Propositions 3 and 4 summarize the role of heterogeneity in primitive risk aversion as a necessary and sufficient condition.

[27] Equilibrium exists if there are at least three traders in a market. Hence, there must be at least six traders in the split market. In addition, for the allocation to be sufficiently affected by the split in at least one group, that group must have at least four traders. One of these four agents is highly risk-averse; if there were only two agents with low risk aversion, the residual supply of each would be effectively determined by one other agent, thus it would be close to inelastic, and so their allocation would be close to their initial endowments, and hence their utility loss from decentralization would be high.

One can further ask: which joint distributions of endowments and risk aversion tend to make the centralized market less efficient than a decentralized market? To identify a joint condition on the primitives, one must account for both the inefficiency due to imperfect diversification of the initial endowment ($\Gamma_i < \mathrm{Id}$) and the fact that the equilibrium aggregate portfolio differs from the efficient one. Proposition 10 in online Appendix D provides two general conclusions: *a market structure gives rise to low welfare compared to other market structures when the highest-price impact agents have the largest need to trade* (i.e., a large (in absolute value) inefficient part of the endowments). Moreover, since $\mathbf{Q}^*$ assigns higher weights to the endowments of less risk-averse traders (cf. (19)), when traders with low risk aversion have large (nondiversifiable parts of) endowments, the difference between the efficient and equilibrium allocations for the traders with low and high risk aversion tends to be the largest. In particular, low risk-averse agents trade too little, and high risk-averse agents trade too much, as in Example 2.

Splitting the market into two disjoint exchanges in Example 2 is an extreme instance of market decentralization. An intermediated market may further improve welfare over a market with disjoint exchanges (Example 6 in online Appendix D). One can show that when intermediation improves welfare over the centralized *one-asset market*, a decrease in the utility of the trader who connects exchanges[28] (relative to an otherwise identical trader who does not intermediate) is a general feature of two-exchange decentralized markets. Intermediation is socially efficient but requires compensation. In the next section, we show that types of intermediation specific to *multiple-asset markets* do not require compensation to be welfare-improving.

D. *Idiosyncratic Risk*

In this section, we show that a change in the market structure may improve diversification of *idiosyncratic risk*, that is, risk whose changes do not affect the risk that is nondiversifiable in equilibrium $\mathbf{Q}^*$. The possibility to improve welfare by changing idiosyncratic risk alone is specific to markets with multiple assets. We compare two decentralized markets, $\mathcal{M}$ and $\mathcal{M}'$, such that $\mathcal{M}'$ is more decentralized than $\mathcal{M}$. To isolate the role that idiosyncratic risk plays in determining welfare, suppose that there is no aggregate risk in either market: $\mathbf{Q}^{*\mathcal{M}} = \mathbf{Q}^{*\mathcal{M}'} = 0$.[29] While then the centralized market is best by Proposition 2(i), Proposition 5 shows that decentralizing a market may increase welfare in the Pareto sense.

PROPOSITION 5 (Market Decentralization Can Be Pareto Improving): *Consider two decentralized markets $\mathcal{M}$, $\mathcal{M}'$ and suppose that $\mathbf{Q}^{*\mathcal{M}} = \mathbf{Q}^{*\mathcal{M}'} = 0$. Let also $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_I$.*

[28] This contrasts our welfare effects with those based on price discrimination, where the objective of the monopolist (here, the connecting trader) typically involves maximization of the payment alone and is not affected by diversification motives.

[29] Consider a subset J of traders who all have the same risk aversion and participation; $(\alpha_i, N(i))$ is independent of i for all $i \in J$. Then, all traders in J submit identical demand schedules; their aggregate (net) demand depends only on the aggregate endowment of J, $Q_J^0 = \sum_{i \in J} q_i^0$; and the aggregate risk $\mathbf{Q}^*$ is independent of the distribution of initial endowments within the class.

(*i*) *If there is one asset, the total welfare is always higher in the more centralized market.*

(*ii*) *If there are at least two assets, $I \geq 5$, and α_{I-1}/α_3 is sufficiently large, then there exist an open set of covariance matrices, a non-empty set of endowments, and two decentralized markets $\mathcal{M}$ and $\mathcal{M}'$ such that $\mathcal{M}'$ is more decentralized than $\mathcal{M}$, and for any covariance Σ in this set, each trader is better off in $\mathcal{M}'$.*

Result (ii) is due to the endogeneity of demand substitutability in decentralized markets: the nonproportionality of price impact in the fundamental covariance Σ implies that demand substitutability, defined by the cross-asset elements of the demand slope

$$S_i \equiv \left(\alpha_i \mathbf{V}_{N(i)} + \Lambda_i\right)^{-1} = \begin{pmatrix} S_j^{11} & S_j^{12} \\ S_j^{21} & S_j^{22} \end{pmatrix}, \tag{24}$$

is endogenous and differs from the payoff substitutability of the assets, as defined by the exogenous covariance Σ. By definition, *demand substitutes* have a positive cross-price elasticity (a higher price of one asset increases the demand for other assets), which corresponds to the negative off-diagonal element of the demand slope. *Payoff substitutes* correspond to positive off-diagonal elements of Σ. In the centralized market, demand substitutability always corresponds to the fundamental substitutability of the assets. This applies in decentralized markets when Λ_i is proportional to $\mathbf{V}_{N(i)}$ (e.g., the case with disjoint exchanges). More generally, decentralizing a market changes how assets substitute in traders' demands. This may increase welfare despite the larger price impact. Example 5 illustrates this point.

Example 5 (Endogenous Demand Substitutability and Welfare): There are two classes of agents, with risk aversions α_i and M_i agents in each class $i = 1, 2$, and two assets with a nonsingular covariance matrix and variances normalized to 1, $\mathbf{V} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. In market $\mathcal{M}$, agents of class 1 trade both assets 1 and 2; agents of class 2 only trade asset 2. Price impacts $\Lambda_1 \in \mathbb{R}^{2\times 2}$, $\Lambda_2 \in \mathbb{R}$ satisfy

$$\Lambda_1 = \left(\begin{pmatrix} 0 & 0 \\ 0 & M_2(\Lambda_2 + \alpha_2)^{-1} \end{pmatrix} + (M_1 - 1)(\Lambda_1 + \alpha_1 \mathbf{V})^{-1}\right)^{-1}, \tag{25}$$

$$\Lambda_2 = \left(\left(\begin{pmatrix} 0 & 0 \\ 0 & (M_2 - 1)(\Lambda_2 + \alpha_2)^{-1} \end{pmatrix} + M_1(\Lambda_1 + \alpha_1 \mathbf{V})^{-1}\right)^{-1}\right)_{22}.$$

Consider a more decentralized market $\mathcal{M}'$, in which some number M of class 1 agents trade only asset 1 and the remaining class members, whom we denote as $1d$, trade both assets; $M_{1d} = M_1 - M$. In terms of participation, $\mathcal{M}$ is characterized by $\{N(1) = \{1, 2\}, N(2) = \{2\}\}$, while $\mathcal{M}'$ is characterized by $\{N'(1 \backslash 1d) = \{1\}$, $N'(2) = \{2\}$, $N'(1d) = \{1, 2\}\}$.

In market $\mathcal{M}'$, agents $1d$ serve as intermediaries who trade with other agents. In this regard, $\mathcal{M}'$ resembles the market structure in Examples 3 and 4, the main differences being that the two assets are not identical and intermediation is not monopolistic. Price impacts $\Lambda_1', \Lambda_2', \Lambda_{1d}$ are characterized by the system

$$\Lambda_1' = \left(\left(\begin{pmatrix}(M-1)(\Lambda_1'+\alpha_1)^{-1} & 0\\ 0 & M_2(\Lambda_2'+\alpha_2)^{-1}\end{pmatrix} + M_{1d}(\Lambda_{1d}+\alpha_1\mathbf{V})^{-1}\right)^{-1}\right)_{11},$$

$$\Lambda_2' = \left(\left(\begin{pmatrix}M(\Lambda_1'+\alpha_1)^{-1} & 0\\ 0 & (M_2-1)(\Lambda_2'+\alpha_2)^{-1}\end{pmatrix} + M_{1d}(\Lambda_{1d}+\alpha_1\mathbf{V})^{-1}\right)^{-1}\right)_{22},$$

$$\Lambda_{1d} = \left(\begin{pmatrix}M(\Lambda_1'+\alpha_1)^{-1} & 0\\ 0 & M_2(\Lambda_2'+\alpha_2)^{-1}\end{pmatrix} + (M_{1d}-1)(\Lambda_{1d}+\alpha_1\mathbf{V})^{-1}\right)^{-1}.$$

Suppose that the heterogeneity in risk aversions is sufficiently large, and $\rho > 0$. There are 6 traders in the market, $M_1 = M_2 = 3$ and $M = 1$, so that $M_{1d} = M_1 - M = 2$.

Class 2 agents as well as agents from class $1\backslash 1d$ are worse off in the more decentralized market $\mathcal{M}'$, regardless of their endowments. Indeed, class 2 agents trade only one asset, and hence Proposition 2(i) implies that $\bar{\Upsilon}_2^{\mathcal{M}} \geq \bar{\Upsilon}_2^{\mathcal{M}'}$.[30] Likewise, since agents $1\backslash 1d$ lose the ability to diversify in the exchange for asset 2, we naturally expect that their price impact increases, and $\bar{\Upsilon}_i^{\mathcal{M}} \geq \bar{\Upsilon}_i^{\mathcal{M}'}$, $i \in \{1\backslash 1d\}$; thus, their utility surplus decreases as well for all endowments. On the other hand, agents of class $1d$ may be better off in $\mathcal{M}'$: by Proposition 2 (ii), it is possible that $\bar{\Upsilon}_{1d}^{\mathcal{M}} \not\geq \bar{\Upsilon}_{1d}^{\mathcal{M}'}$ because a change in the market structure affects how trades of different assets substitute in creating utility, i.e., $(\bar{\Upsilon}_{1d}^{\mathcal{M}'})_{1,2}$. This can offset the within-exchange effect and increase equilibrium utility for some initial endowments, as we show next.

By assumption, two assets are payoff substitutes ($\rho > 0$), and hence they are complements in demand slopes and the utility surplus matrix; that is, $(\bar{\Upsilon}_{1d}^{\mathcal{M}'})_{1,2} < 0$. Crucially, this surplus complementarity between assets 1 and 2 is endogenous, since it depends on price impact, and decreases when market $\mathcal{M}$ becomes more

[30] Since they can only trade one asset, both Λ_2 and $\mathbf{V}_{N(2)}$ are scalar, and hence Λ_2 is always proportional to $\mathbf{V}_{N(2)}$.

decentralized: $(\bar{\Upsilon}_{1d}^{\mathcal{M}'})_{1,2} < (\bar{\Upsilon}_{1d}^{\mathcal{M}})_{1,2}$. By equation (23), this lower complementarity can increase utility if and only if $\bar{\Upsilon}_{1d}^{\mathcal{M}} \not\geq \bar{\Upsilon}_{1d}^{\mathcal{M}'}$, which is equivalent to

$$|\left(\bar{\Upsilon}_{1d}^{\mathcal{M}}\right)_{1,2} - \left(\bar{\Upsilon}_{1d}^{\mathcal{M}'}\right)_{1,2}|^2 > \prod_{i=1}^{2}\left(\left(\bar{\Upsilon}_{1d}^{\mathcal{M}}\right)_{i,i} - \left(\bar{\Upsilon}_{1d}^{\mathcal{M}'}\right)_{i,i}\right), \tag{26}$$

using that $\bar{\Upsilon}_{1d}^{\mathcal{M}} - \bar{\Upsilon}_{1d}^{\mathcal{M}'}$ is not positive semidefinite if and only if its determinant is negative. That is, the utility gain from the weaker complementarity (LHS) must exceed the loss from the higher price impact in each market (RHS of (26)). If $\bar{\Upsilon}_{1d}^{\mathcal{M}} - \bar{\Upsilon}_{1d}^{\mathcal{M}'}$ is sufficiently large, the total welfare increases despite the lower surpluses of classes 1 and 2. We used that equilibrium utility (23) allows attributing welfare effects of market decentralization to changes in price impact of different assets and exchanges.

When can one expect the welfare effect of the change in idiosyncratic risk to be significant? Assets must be correlated imperfectly and sufficiently strongly. It is clear that assets cannot be independent: if correlation ρ is close to 0, all price impacts and surplus matrices are almost independent (the LHS of (26) is close to 0), implying that all agents are worse off in the more decentralized market. They cannot be perfectly correlated either: if the correlation ρ is close to 1, both surplus matrices $\bar{\Upsilon}_{1d}^{\mathcal{M}}$, $\bar{\Upsilon}_{1d}^{\mathcal{M}'}$ and their difference are proportional to the (almost singular) matrix $\mathbf{V}$, and thus the difference between the left-hand side and the right-hand side of (26) is close to 0.

Hence, the possibility of a welfare increase by reduction of idiosyncratic risk in Proposition 5 is a multi-asset effect. In one-asset markets, while the asset traded in different exchanges can be an imperfect demand substitute, its trades are perfect substitutes in the utility surplus Υ_i.[31] Since with constant aggregate risk $\mathbf{Q}^*$, $(\mathbf{Q}^*_{N(i)} - \alpha_i \mathbf{V}_{N(i)} q_i^0) \equiv X_i$ is independent of market structure, Theorem 2 implies that the risk premium benefit $X_i^T S_i X_i$ in the utility surplus is monotone decreasing in market decentralization (demand slope S_i is always smaller in the more decentralized market). Thus, for the utility to increase, the cost $0.5\alpha_i X_i^T S_i \mathbf{V}_{N(i)} S_i X_i$ must decrease sufficiently to offset the decrease in $X_i^T S_i X_i$. However, in one-asset markets, the second part is proportional to $\mathbf{V}$, and hence singular. One can show that the total welfare gain from decentralization is inverse U-shaped, and for the total welfare surplus to be positive, ρ must be sufficiently different from 0 and 1. Then, there exists an open set of initial endowments such that agents of class $1d$ are strictly better off in $\mathcal{M}'$, while other agents have zero endowments and the same utility in $\mathcal{M}$ and $\mathcal{M}'$. That is, $\mathcal{M}'$ weakly Pareto dominates $\mathcal{M}$.

In the more decentralized market $\mathcal{M}'$, some agents of class 1 specialize in trading asset 1 and refrain from trading asset 2. While such specialization unambiguously increases price impact (Theorem 2), Example 5 shows that it can be Pareto improving, even among equally risk-averse agents. The gains from trade in the more decentralized market structure come from the effect of specialization to lower the demand substitutability of imperfectly correlated assets, which *decreases the utility*

[31] The (welfare-improving) decrease in surplus substitutability $\Upsilon_{n,m}$ is exactly offset by the liquidity decrease $\Upsilon_{n,n}$ within each exchange n.

cost of holding correlated risky assets $\Upsilon_{n,m}$, $m \neq n$. In fact, the same assets can turn from demand and utility substitutes into complements (i.e., in the sense of S_i and Υ_i, respectively). Changing who trades which assets without necessarily changing who trades with whom allows *reducing idiosyncratic risk without affecting aggregate risk*.

With strategic traders and heterogeneous risk preferences, both centralized and decentralized markets can generally be Pareto efficient (i.e., no other market structure would make some agents better off, without making others worse off), hence our focus on total welfare. This section shows that regulation that aims to make an existing market structure less decentralized, if applied without discretion, might have unintended consequences for utility of all agents.

IV. Conclusions

The results in this paper recognize the potential for decentralized markets to increase welfare compared to the centralized market. This holds even with homogeneous assets and even if locating counterparties has a cost (which we have taken to be 0 throughout). We have shown that when trading is noncompetitive, the centralized market is efficient for all endowment distributions only if risk preferences of all market participants are sufficiently symmetric. Decentralized market structures may strictly dominate in total welfare sense when risk preferences and need to trade are sufficiently heterogeneous for some market participants. We identify two ways in which decentralized trading can enhance the role of markets in allocating risk: it may reallocate more risk to less risk-averse traders despite the larger price impact, by enabling agents to trade distinct components of aggregate risk. Additionally, by changing how the trade of correlated assets substitutes in creating equilibrium utility, it may allow traders to reduce idiosyncratic risk, even without affecting aggregate risk.

We have taken the exchanges and available assets in decentralized markets as exogenous. Nevertheless, our results imply that in general, allowing agents to choose with whom they want to trade and which assets will not result in the centralized market—some traders would prefer to trade separately. Our analysis also suggests that the study of the endogenous formation of exchanges in decentralized markets, with respect to welfare or other objectives, should not be separate from security design. In fact, the endogenously heterogeneous demand substitutability for identical assets traded in decentralized markets implies the existence of profit and efficiency opportunities from security design as well as specialization in trading certain assets that are not available in centralized markets.

Our results suggest a rich theory of the ways in which intermediation—a decentralized market phenomenon—can improve efficiency and of the forms such intermediation can take. We show that the introduction of strategic intermediaries may mitigate both inefficiency due to imperfect diversification of idiosyncratic risk and inefficient aggregate risk.

The centralized market theory of asymmetric information has provided arguments as to why private information (e.g., adverse selection) or more information (e.g., information revealed through prices) may create incentives for some traders to trade in a separate market. When these incentives are also associated with a welfare increase, those results suggest additional reasons for decentralization when trade is motivated by not only diversification but also asymmetric information (e.g., Rostek

and Weretka 2015b; Babus and Kondor 2016; the lemon markets model by Akerlof 1970 could also be interpreted this way). This paper shows that arguments based on risk sharing and diversification can be made for decentralization of trade, even in the absence of information-related reasons.

One might wonder whether the uniform-price mechanism is significant to the main conclusions. The starting point for identifying the possibility of welfare improvements is the inefficiency of (centralized or decentralized market) allocation due to price impact. Based on the results in the theory of games with (two-sided) private information, one expects inefficiency to be present in equilibrium and the welfare gains from market decentralization to exist for other pricing mechanisms. Fundamentally, these gains stem from the possibilities that decentralized trading creates for alignment of agents' risk preferences and their equilibrium risk exposure via the market structure. One might also wonder if our conclusions based on a static model are useful given that most markets are dynamic. Dynamic trading changes incentives and motivates a separate study of welfare effects. In a dynamic model, traders have price impact in every trading round. The centralized market inefficiency and all the decentralized market effects that we identify in the static model are present in all rounds. With gains to diversification renewed through endowment shocks, the static effect we identify will be first-order. Our results suggest the potential for welfare improvements in markets in the cross section (their market structure) as well as in time series.

Appendix

A. *Equilibrium Characterization*

PROOF OF PROPOSITION 1:

(i), (ii) That the equilibrium characterized by schedules and price impacts $\{q_i(\cdot, \Lambda_i), \Lambda_i\}_i$ is equivalent to a Bayesian Nash Equilibrium in a decentralized market follows from Lemma 1 in Rostek and Weretka (2015a).

(iii) For uniqueness, diagonalize Σ by multiplying (6) from the left and right by $\Sigma^{-1/2}$ and denote $\tilde{\Lambda}_i = \Sigma^{-1/2}\Lambda_i\Sigma^{-1/2}$,

$$\tilde{\Lambda}_i = \left(\sum_{j\neq i}(\alpha_j \mathrm{Id} + \tilde{\Lambda}_j)^{-1}\right)^{-1}, \quad i \in I. \tag{27}$$

One expects that any solution to this equation is of the form $\tilde{\Lambda}_i = \beta_i\alpha_i\mathrm{Id}$ for some $\beta_i > 0$, $i \in I$, and consequently $\Lambda_i = \beta_i\alpha_i\Sigma$. Lemma C.4 in online Appendix C shows that this is the case. The analytic characterization of equilibrium follows by Lemma D.1 in online Appendix D.

Part (iv) follows directly from the harmonic mean condition: defining $B = \sum_i(\Lambda_i + \alpha_i\Sigma)^{-1}$, and noting that by (iii), $\Lambda_i = \beta_i\alpha_i\Sigma$ for some β_i, we get $B = b\Sigma^{-1}$ for some $b > 0$, and (ii) implies $\Lambda_i = (B - (\Lambda_i + \alpha_i\Sigma)^{-1})^{-1}$, which is equivalent to $\beta_i\alpha_i = (b - (\beta_i\alpha_i + \alpha_i)^{-1})^{-1}$. This is a quadratic equation for β_i, and it has only one positive solution given by (7). Substituting β_i into $b = \sum_i(\beta_i\alpha_i + \alpha_i)^{-1}$, we get the required equation for b. ∎

To prove Theorem 1, we will need several auxiliary results. Let $\mathcal{S}^I$ be the set of I-tuples $\{\Lambda_i\}_i$ of positive semidefinite matrices with $\Lambda_i \in \mathbb{R}^{N(i)\times N(i)}$. On this set, we introduce a partial order: $\{\Lambda_i\}_i \leq \{\Lambda_i'\}_i$ for a pair of tuples $\{\Lambda_i\}_i$, $\{\Lambda_i'\}_i$ if $\Lambda_i \leq \Lambda_i'$ for all $i \in I$. Recalling that the negative X_i of the slope of i's demand and i's price impact are linked through $X_i \equiv (\alpha_i \mathbf{V}_{N(i)} + \Lambda_i)^{-1}$, we can rewrite the fixed point condition (15) as a fixed point condition for demand slopes as follows. Define the map $G = \{G_i\}_i : \mathcal{S}^I \to \mathcal{S}^I$ via

$$G_i(\{X_i\}_i) = \left(\left(\left(\sum_{j\neq i} \bar{X}_j\right)^{-1}\right)_{N(i)} + \alpha_i \mathbf{V}_{N(i)}\right)^{-1}, \quad i \in I. \tag{28}$$

Essentially, the decentralized market model can be seen as a game in which agents choose their demand slopes. Let us denote by $G^n(\{\Lambda_i\}_i)$ the nth iteration of the best response map. Standard properties of the positive semidefinite order imply that G is monotone increasing in $\{X_i\}_i$.

LEMMA 1: *Map G is monotone increasing on $\mathcal{S}^I$.*

Let $F = \{F_i\}_i : \mathcal{S}^I \to \mathcal{S}^I$ be the map defined by the right-hand side of (15). By construction, maps F and G are simple transformations of each other. It is more convenient analytically, however, to work directly with map F; consequently, all proofs that follow use this map. Passing from F to G is then straightforward. The following result then follows from Theorem 1.

LEMMA 2: *A tuple of linear demand schedules with slopes $\{X_i\}_i$ is an equilibrium if and only if it is a fixed point of the best response map. That is, $\{X_i\}_i = G(\{X_i\}_i)$.*

PROPOSITION 6 (Monotone Convergence): *Pick an arbitrary starting tuple $\{X_i^0\}_i$ such that $\{X_i^0\}_i \leq G(\{X_i^0\}_i)$ ($\{X_i^0\}_i \geq G(\{X_i^0\}_i)$). Then, iteration $G^n(\{X_i^0\}_i)$ is monotone increasing (decreasing) in n and converges to an equilibrium tuple.*

Let $D = \operatorname{diag}(z)$, $z \in \mathbb{R}^N$ be a diagonal matrix. Multiplication of (15) by $D_{N(i)}$ from the left and right gives the following scale invariance property of price impacts.

LEMMA 3: *Let $\mathbf{V}' = D\mathbf{V}D$. Then, the map given by $\{\Lambda_i\}_i \to \{D_{N(i)}\Lambda_i D_{N(i)}\}_i$ defines a one-to-one correspondence between equilibria in markets defined by $\mathbf{V}$ and $\mathbf{V}'$.*

PROOF OF THEOREM 1:

(i), (ii) Treating assets traded in different exchanges as distinct assets and lifting before aggregation gives the first-order condition (4) and the system of I price impact harmonic mean equations. Equation (4) gives demand function (5) for each trader i. Given the lifting procedure, that the equilibrium characterized by schedules and price impacts $\{q_i(\cdot, \Lambda_i), \Lambda_i\}_i$ is equivalent to a Bayesian Nash Equilibrium in a decentralized market follows from Lemma 1 in Rostek and Weretka (2015a).

(iii) Existence of equilibria when $\mathbf{V}_{N(i)}$ is nonsingular for any i follows from Proposition 6. For the general case, let $\mathbf{V}^\varepsilon \equiv \mathbf{V} + \varepsilon \mathrm{Id}$ and let F^ε be the corresponding map. By Proposition 6, for any $\varepsilon > 0$, there exists an equilibrium $\{\Lambda_i^\varepsilon\}_i$ corresponding

to $\mathbf{V}^{\varepsilon}$. Pick a sequence $\varepsilon_k = 1/k$ and an equilibrium $\{\Lambda_i^{\varepsilon_k}\}_i$. Since F^{ε} is monotone increasing in ε, we have

$$\{\Lambda_i^{\varepsilon_k}\}_i = F^{\varepsilon_k}(\{\Lambda_i^{\varepsilon_k}\}_i) \geq F^{\varepsilon_{k+1}}(\{\Lambda_i^{\varepsilon_k}\}_i),$$

and hence by Proposition 6, there exists an equilibrium $\{\Lambda_i^{\varepsilon_{k+1}}\}_i \leq \{\Lambda_i^{\varepsilon_k}\}_i$. Thus, we can construct a monotone decreasing sequence $\{\Lambda_i^{\varepsilon_k}\}_i$ that converges to an equilibrium corresponding to $\varepsilon = 0$.

To prove generic determinacy, let, for each i, $X_i \equiv \Lambda_i + \alpha_i \mathbf{V}_{N(i)}$, and define a map $\Phi : \mathcal{S}^I \to \mathcal{S}^I$ via

$$\Phi_i(\{X_j\}_j) = X_i - \left(\left(\sum_{j \neq i} X_j^{-1}\right)^{-1}\right)_{N(i)}.$$

The equilibrium equation can be written as $\Phi(\{X_j\}_j) = \{\alpha_j \mathbf{V}_{N(j)}\}_j$. Let Ψ^I be the image of the map $\mathbf{V} \to \{\alpha_i \mathbf{V}_{N(i)}\}_i$ defined on the set of positive semidefinite matrices.[32] Let Θ^I be the subset of $\mathcal{S}^I$ such that for any $\{X_j\}_j \in \Theta^I$, we have that $\Phi(\{X_j\}_j) = \{\alpha_j \mathbf{V}_{N(j)}\}_j$ for some positive definite matrix $\mathbf{V}$. Subset Θ^I is an algebraic variety, and therefore can be represented as a finite union of irreducible algebraic sets that are smooth manifolds. The same is true for Ψ^I. By Sard's Theorem, almost every $\{\alpha_i \mathbf{V}_{N(i)}\}_i \in \Psi^I$ has a regular preimage under Φ; that is, equilibria are determinate for generic covariance matrices. The finiteness of the set of equilibria follows by the standard compactness arguments and the fact that all equilibria belong to a compact set.

(iv) For the positive semidefiniteness of price impacts, observe that $\Lambda_i + \alpha_i \mathbf{V}_{N(i)}$ must be positive semidefinite for any trader i in equilibrium. Suppose otherwise. The utility of agent i who acquires portfolio $y \in \mathbb{R}^{N(i)}$ is given by $(q_i^0 + y)^T d - y^T(p^* + \Lambda_i y) - 0.5\,(q_i^0 + y)^T \alpha_i \mathbf{V}_{N(i)}(q_i^0 + y)$, where p^* is the equilibrium price vector if agent i does not trade. If $\Lambda_i + \alpha_i \mathbf{V}_{N(i)}$ were not positive semidefinite, then the agent could attain infinite utility; that is, if there is a y such that $y^T(\Lambda_i + \alpha_i \mathbf{V}_{N(i)})y < 0$, then buying an infinite amount of portfolio y gives an infinite utility. If $\Lambda_i + \alpha_i \mathbf{V}_{N(i)}$ is positive semidefinite for any i, then so is Λ_i by (15).

Let now $C(\mathbf{V}) = \mathrm{diag}(\{\mathbf{V}_{nn}^{-1/2}\})\mathbf{V}\mathrm{diag}(\{\mathbf{V}_{nn}^{-1/2}\})$ be the correlation matrix of the assets. For any exchange n and agent i, define $\alpha_{i,**} \equiv \alpha_i \min(\mathrm{eig}(C(\mathbf{V})_{N(i)}))$ and $\alpha_i^{**} \equiv \alpha_i \max(\mathrm{eig}(C(\mathbf{V})_{N(i)}))$; $\alpha_{i,**}$ and α_i^{**} can be interpreted as the bounds on the assets' effective riskiness (see Section IIID). For any exchange n, define two constants $\lambda_{**}(n) \equiv \min_{i \in I(n)} \alpha_{i,**}$ and $\lambda^{**}(n) \equiv \max_{i \in I(n)} \alpha_i^{**}$. Further, let $\{\Lambda_{i,\min}^0\}_i = \left\{\mathrm{diag}\left(\frac{\lambda_{**}(n)\mathbf{V}_{nn}}{I(n)-2}\right)_{N(i)}\right\}_i$ and $\{\Lambda_{i,\max}^0\}_i = \left\{\mathrm{diag}\left(\frac{\lambda^{**}(n)\mathbf{V}_{nn}}{I(n)-2}\right)_{N(i)}\right\}_i$ and

$$\{X_{i,\min}^0\}_i = \{(\alpha_i \mathbf{V}_{N(i)} + \Lambda_{i,\max}^0)^{-1}\}_i, \quad \{X_{i,\max}^0\}_i = \{(\alpha_i \mathbf{V}_{N(i)} + \Lambda_{i,\min}^0)^{-1}\}_i.$$

[32] Note that if some of the assets are replicas of each other, the covariance matrix $\mathbf{V}$ belongs to a subspace of $\mathbb{R}^{N \times N}$. In this case, we apply Sard's Theorem to the map into this subspace with respect to the induced Lebesgue measure.

A direct calculation implies that

$$\{X^0_{i,\min}\}_i \le G(\{X^0_{i,\min}\}_i) \quad \text{and} \quad \{X^0_{i,\max}\}_i \ge G(\{X^0_{i,\max}\}_i),$$

and, similarly,

$$\{\Lambda^0_{i,\min}\}_i \le F(\{\Lambda^0_{i,\min}\}_i) \quad \text{and} \quad \{\Lambda^0_{i,\max}\}_i \ge F(\{\Lambda^0_{i,\max}\}_i).$$

We now construct the minimal and maximal equilibria by the explicit iterative procedure described in Proposition 6. To this end, recursively define two sequences $\{\Lambda^k_{i,\min}\}_i \in \mathcal{S}^I$ and $\{\Lambda^k_{i,\max}\}_i \in \mathcal{S}^I$, $k \ge 1$ via $\{\Lambda^k_{i,\min}\}_i \equiv F(\{\Lambda^{k-1}_{i,\min}\}_i)$ and $\{\Lambda^k_{i,\max}\}_i \equiv F(\{\Lambda^{k-1}_{i,\max}\}_i)$. By Proposition 6, the sequence $\{\Lambda^k_{i,\min}\}_i$, $k \ge 0$, is monotone increasing, whereas $\{\Lambda^k_{i,\max}\}_i$, $k \ge 0$, is monotone decreasing; these sequences converge to equilibria (the fixed points of map F) that we denote by $\{\Lambda_{i,\min}\}_i$ and $\{\Lambda_{i,\max}\}_i$, respectively. The corresponding demand slopes are determined via $\{X_{i,\min}\}_i = \{(\alpha_i \mathbf{V}_{N(i)} + \Lambda_{i,\max})^{-1}\}_i$, $\{X_{i,\max}\}_i = \{(\alpha_i \mathbf{V}_{N(i)} + \Lambda_{i,\min})^{-1}\}_i$.

Pick an arbitrary equilibrium $\{\Lambda_i\}_i$. Then, for all $i \in I$,

$$\Lambda_i = \left(\left(\sum_{j\neq i}(\alpha_j \mathbf{V} + \bar{\Lambda}_j)^{-1}\right)^{-1}\right)_{N(i)} \le \left(\left(\sum_{j\neq i}(\alpha^{**}_j \mathrm{Id}_{N(j)} + \bar{\Lambda}_j)^{-1}\right)^{-1}\right)_{N(i)}. \tag{29}$$

Let F_A be the map corresponding to the right-hand side of (29). Then, iterating F_A and using Proposition 6, arrive at the conclusion that F_A has a fixed point $\{\Lambda^{**}_i\}_i$ satisfying $\{\Lambda_i\}_i \le \{\Lambda^{**}_i\}_i$. By Lemma C.4 in the online Appendix, this is the unique diagonal fixed point. Then, Λ^{**}_i is diagonal, and for any exchange n, the scalar price impacts $\{(\Lambda^{**}_i)_{nn}\}_i$ coincide with price impacts in a centralized exchange for a single asset with variance 1 and risk aversion α^{**}_i. The same iteration argument as above implies that these price impacts are monotone increasing in α^{**}_i, and therefore, satisfy

$$\Lambda^{**}_i \le \frac{\lambda^{**}(n)}{I(n)-2}\mathrm{Id}_{N(i)}, \quad i \in I.$$

Hence, by the monotonicity of map F, $\{\Lambda_i\}_i = F^n(\{\Lambda_i\}_i) \le F^n(\{\Lambda^0_{i,\max}\}_i) \to \{\Lambda_{i,\max}\}_i$. Similarly,

$$\Lambda_i = \left(\left(\sum_{j\neq i}(\alpha_j \mathbf{V} + \bar{\Lambda}_j)^{-1}\right)^{-1}\right)_{N(i)} \ge \left(\left(\sum_{j\neq i}(\alpha_{j,**} \mathrm{Id}_{N(j)} + \bar{\Lambda}_j)^{-1}\right)^{-1}\right)_{N(i)}.$$

The same argument as above implies that

$$\Lambda_{i**} \ge \frac{\lambda_{**}(n)}{I(n)-2}\mathrm{Id}_{N(i)}, \quad i \in I,$$

and the same iterative procedure completes the proof: $\{\Lambda_i\}_i = F^n(\{\Lambda_i\}_i) \ge F^n(\{\Lambda^0_{i,\min}\}_i) \to \{\Lambda_{i,\min}\}_i$. ∎

B. *Comparative Statics and Welfare*

PROOF OF THEOREM 2:

(i) Consider two markets which differ only in participation $\{N(i)\}_i$ and $\{N'(i)\}_i$, $N(i)' \supseteq N(i)$ for all $i \in I$. Pick an equilibrium $\{\Lambda_i\}_i$ corresponding to $\{N'(i)\}_i$. By Lemma E.2 in the online Appendix,

$$(\Lambda_i)_{N(i)} = \left(\left(\sum_{j\neq i}(\alpha_j \bar{\mathcal{V}}_{N'(j)} + \bar{\Lambda}_j)^{-1}\right)^{-1}\right)_{N(i)} \leq \left(\left(\sum_{j\neq i}(\alpha_j \bar{\mathbf{V}}_{N(j)} + (\bar{\Lambda}_j)_{N(j)})^{-1}\right)^{-1}\right)_{N(i)}$$

for all $i \in I$. Therefore, by Proposition 6, there exists an equilibrium $\{\hat{\Lambda}_i\}_i$ corresponding to participation $\{N(i)\}_i$ and satisfying $\{\hat{\Lambda}_i\}_i \geq \{(\Lambda_i)_{N(i)}\}_i$; the claim follows.

(ii) Fix a parameter α and let us rewrite the equilibrium equation as $\{\Lambda_j\}_j = F(\{\Lambda_j\}_j, \alpha)$. By definition, for both $\alpha = M_i$ and $\alpha = \alpha_j$, map F is monotone increasing in α in the sense of the positive semidefinite partial order.

Fix $\alpha_1 \leq \alpha_2$ and let $\{\Lambda_j(\alpha_1)\}_j$ be an equilibrium. Then

$$\{\Lambda_j(\alpha_1)\}_j = F(\{\Lambda_j(\alpha_1)\}_j, \alpha_1) \leq F(\{\bar{\Lambda}_j(\alpha_1)\}_j, \alpha_2).$$

By Proposition 6, there exists an equilibrium $\{\Lambda_j(\alpha_2)\}_j$ satisfying $\{\Lambda_j(\alpha_2)\}_j \geq \{\Lambda_j(\alpha_1)\}_j$.

The next claims follow by a similar argument using Theorem 24 in Anderson and Duffin (1969).

Finally, part (iii) follows from Lemma D.3 in the online Appendix. ∎

PROOF OF PROPOSITION 2:

The expression for the surplus matrix follows by direct calculation: denoting by $\cdot$ the inner product in $\mathbb{R}^N$, we observe that the equilibrium utility is given by

(30)

$$\begin{aligned}
&\bar{q}_i^0 \cdot d - \bar{q}_i \cdot (p - d) - 0.5(\bar{q}_i^0 + \bar{q}_i) \cdot \alpha_i \mathbf{V}(\bar{q}_i^0 + \bar{q}_i)\\
&= q_i^0 \cdot d + \bar{q}_i \cdot (\mathbf{Q}^* - \alpha_i \mathbf{V}\bar{q}_i^0 + \alpha_i \mathbf{V}\bar{q}_i^0) - 0.5(\bar{q}_i^0 + \bar{q}_i) \cdot \alpha_i \mathbf{V}(\bar{q}_i^0 + \bar{q}_i)\\
&= q_i^0 \cdot d - 0.5\,\bar{q}_i^0 \cdot \alpha_i \mathbf{V}\bar{q}_i^0 + \bar{q}_i \cdot (\mathbf{Q}^* - \alpha_i \mathbf{V}\bar{q}_i^0) - 0.5\,\bar{q}_i \cdot \alpha_i \mathbf{V}\bar{q}_i\\
&= q_i^0 \cdot d - 0.5\,\bar{q}_i^0 \cdot \alpha_i \mathbf{V}\bar{q}_i^0 + (\mathbf{Q}^* - \alpha_i \mathbf{V}\bar{q}_i^0) \cdot (\bar{\Lambda}_i + \alpha_i \bar{\mathbf{V}}_{N(i)})^{-1}(\mathbf{Q}^* - \alpha_i \mathbf{V}\bar{q}_i^0)\\
&\quad - 0.5(\mathbf{Q}^* - \alpha_i \mathbf{V}\bar{q}_i^0)(\bar{\Lambda}_i + \alpha_i \bar{\mathbf{V}}_{N(i)})^{-1}\alpha_i \mathbf{V}(\bar{\Lambda}_i + \alpha_i \bar{\mathbf{V}}_{N(i)})^{-1}(\mathbf{Q}^* - \alpha_i \mathbf{V}\bar{q}_i^0),
\end{aligned}$$

and the claim follows because

$$\bar{\Upsilon}_i(\Lambda) = (\bar{\Lambda}_i + \alpha_i \bar{\mathbf{V}}_{N(i)})^{-1} - 0.5(\bar{\Lambda}_i + \alpha_i \bar{\mathbf{V}}_{N(i)})^{-1}\alpha_i \mathbf{V}(\bar{\Lambda}_i + \alpha_i \bar{\mathbf{V}}_{N(i)})^{-1}. \quad (31)$$

(i) Suppose that we change a market structure $\mathcal{M}$ to a more decentralized market structure $\mathcal{M}'$. Let Λ_i, Λ_i' be the corresponding price impacts of agent i, and denote by X and X' the slopes of the demand schedules in the two market structures. It follows from Theorem 2 that $\Lambda_i' \geq (\Lambda_i)_{N'(i)}$, and therefore by Lemma E.2 in the online Appendix,

$$(32)\quad X = (\alpha_i \bar{\mathbf{V}}_{N(i)} + \bar{\Lambda}_i)^{-1} \geq (\alpha_i \bar{\mathbf{V}}_{N'(i)} + (\bar{\Lambda}_i)_{N'(i)})^{-1} \geq (\alpha_i \bar{\mathbf{V}}_{N'(i)} + \bar{\Lambda}_i')^{-1} = X'.$$

Using formula (32), we can see that it suffices to show that the inequality $X - 0.5\,\alpha_i X\mathbf{V}X \geq X' - 0.5\,\alpha_i X'\mathbf{V}X'$ holds if X is proportional to $\mathbf{V}$ and $X \geq X'$. Let $X = \beta\mathbf{V}^{-1}$ and let $Y = \mathbf{V}^{1/2}X'\mathbf{V}^{1/2}$. Note that $X \leq (\alpha_i \mathbf{V})^{-1}$ and hence $\beta \leq \alpha_i^{-1}$. Then, we have $Y \leq \beta\mathrm{Id}$. Multiplying the required inequality from both sides by $\mathbf{V}^{1/2}$ yields

$$X - 0.5\,\alpha_i X\mathbf{V}X \geq X' - 0.5\,\alpha_i X'\mathbf{V}X' \Leftrightarrow (\beta - 0.5\,\alpha_i\beta^2)\mathrm{Id} \geq Y - 0.5\,\alpha_i Y^2,$$

and the claim follows because $f(x) = x - 0.5\alpha_i x^2$ is monotone increasing on $[0, \alpha_i^{-1}]$.

Part (ii) follows because by the Löwner Theorem (Donoghue 1974), the function $f(x)$ is not matrix monotone. ∎

PROOF OF PROPOSITION 3:

There exists a map (a matrix) $\mathcal{E}^{\mathcal{M}}$ such that the equilibrium allocation is given by

$$(q_i + q_i^0)_i = (\mathcal{E}^{\mathcal{M}}(q_i^0)_i)_i,$$

where $\mathcal{E}^{\mathcal{M}}$ is a function of equilibrium price impacts Λ_i, and can be written as such explicitly using Corollary 2. The efficiency of a given market structure in allocating risks is then encoded in the way the matrix $\mathcal{E}^{\mathcal{M}}$ "redistributes" endowments. For instance, in the centralized competitive market with a single asset,

$$\mathcal{E}^{\mathcal{M}}(q_i^0)_i = (\alpha_i^{-1})_i q^{**}.$$

That is, the image of the matrix $\mathcal{E}^{\mathcal{M}}$ is one-dimensional and coincides with the span of the efficient allocation (α_i^{-1}). Hence, $\mathcal{E}^{\mathcal{M}}$ is a projection onto the efficient allocation, and it has only two eigenvalues, 1 and 0. Thus, decomposing any vector of endowments into the efficient and the inefficient part,

$$(q_i^0)_i = (\alpha_i^{-1})_i q^{**} + ((q_i^0)_i - q^{**}(\alpha_i^{-1})_i),$$

the matrix $\mathcal{E}^{\mathcal{M}}$ keeps the efficient part unchanged and completely eliminates the inefficient part.

Denote by $\bar{\mathcal{S}}_i \equiv (\bar{\Lambda}_i + \alpha_i \bar{\mathbf{V}}_{N(i)})^{-1}$ the lifted slope of agent i's demand. We have

$$(33)\qquad (\mathcal{E}^{\mathcal{M}}(q_i^0)_i)_j = q_j^0 + \bar{\mathcal{S}}_i\Big(B^{-1}\sum_i \bar{\mathcal{S}}_i\alpha_i\mathbf{V}\bar{q}_i^0 - \alpha_i\mathbf{V}\bar{q}_j^0\Big).$$

Let $\mathcal{D} = \text{diag}(\alpha_i \mathbf{V})_i$ and $\mathcal{B}^{\mathcal{M}} \equiv (\bar{\mathcal{S}}_i B^{-1} \bar{\mathcal{S}}_j)_{i,j=1}^{I}$ and $\mathcal{P}^{\mathcal{M}} = \text{diag}(\bar{\mathcal{S}}_i)$. Then, we have

$$\mathcal{E}^{\mathcal{M}} = \text{Id} + \left(\mathcal{B}^{\mathcal{M}} - \mathcal{P}^{\mathcal{M}}\right)\mathcal{D},$$

and equilibrium total welfare loss can be rewritten as $((q_i^0))^T \mathcal{A}^{\mathcal{M}} (q_i^0)$, where

$$\mathcal{A}^{\mathcal{M}} = \left(\mathcal{E}^{\mathcal{M}}\right)^T \mathcal{D}\mathcal{E}^{\mathcal{M}} = \left(\text{Id} + \mathcal{D}\left(\mathcal{B}^{\mathcal{M}} - \mathcal{P}^{\mathcal{M}}\right)\right)\mathcal{D}\left(\text{Id} + \left(\mathcal{B}^{\mathcal{M}} - \mathcal{P}^{\mathcal{M}}\right)\mathcal{D}\right).$$

Suppose first that all risk aversions are symmetric. For equally risk-averse traders it is efficient to hold equal shares of the total market portfolio; hence, the set of efficient allocations coincides with vectors (q_i^0) for which q_i^0 is independent of i. Denote the subspace of such vectors by $\mathcal{X} \subset \mathbb{R}^{K \times N}$. Then, the matrix $\mathcal{A}^{\mathcal{M}}$ keeps that subspace invariant in the sense that $\mathcal{A}^{\mathcal{M}} x = x$ for all $x \in \mathcal{X}$, and this holds independent of the market structure $\mathcal{M}$. Indeed, this is the case because both $\mathcal{E}^{\mathcal{M}} = (\mathcal{E}^{\mathcal{M}})^T$ and $\mathcal{D}$ keep this subspace invariant, and hence so does $\mathcal{A}^{\mathcal{M}} = (\mathcal{E}^{\mathcal{M}})^T \mathcal{D}\mathcal{E}^{\mathcal{M}}$. The orthogonal complement $\mathcal{X}^{\perp}$ of $\mathcal{X}$ in $\mathbb{R}^{K \times N}$ coincides with the set of initial endowment allocations for which the aggregate endowment is 0. Our goal is to show that the utility loss in a decentralized market is always higher; that is, that $\mathcal{A}^{\mathcal{M}} \geq \mathcal{A}^{*}$. Denote by $\mathcal{R}$ the orthogonal projection on $\mathcal{X}^{\perp}$. Then, by Corollary 1, the centralized market matrix * on that subspace coincides with $\left(\frac{\beta}{1+\beta}\right)^2 \mathcal{D}$. That is, $\mathcal{R}\mathcal{A}^{*}\mathcal{R} = \left(\frac{\beta}{1+\beta}\right)^2 \mathcal{R}\mathcal{D}\mathcal{R}$. Since the matrices $\mathcal{A}^{\mathcal{M}}$ and $\mathcal{A}^{*}$ both keep $\mathcal{X}$ invariant, it suffices to show that

$$\mathcal{R}\mathcal{A}^{\mathcal{M}}\mathcal{R} \geq \mathcal{R}\mathcal{A}^{*}\mathcal{R} = \left(\frac{\beta}{1+\beta}\right)^2 \mathcal{R}\mathcal{D}\mathcal{R}.$$

Denote $\mathcal{B} = \mathcal{D} + \mathcal{D}(\mathcal{B}^{\mathcal{M}} - \mathcal{P}^{\mathcal{M}})$. Then, by direct calculation, $\mathcal{A}^{\mathcal{M}} = \mathcal{B}\mathcal{D}^{-1}\mathcal{B}$ and hence the required inequality takes the form

$$\mathcal{R}\mathcal{B}\mathcal{D}^{-1}\mathcal{B}\mathcal{R} \geq \left(\frac{\beta}{1+\beta}\right)^2 \mathcal{R}\mathcal{D}\mathcal{R}.$$

Using the fact that $\mathcal{D}$ and $\mathcal{R}$ and $\mathcal{D}$ and $\mathcal{D}$ commute and multiplying the required inequality by $\mathcal{D}^{-1/2}$ from both sides, we conclude that we need to show that

$$(\mathcal{R}\mathcal{D}^{-1/2}\mathcal{B}\mathcal{D}^{-1/2}\mathcal{R})^2 \geq \left(\frac{\beta}{1+\beta}\right)^2 \mathcal{R},$$

which is equivalent to $\mathcal{R}\mathcal{D}^{-1/2}\mathcal{B}\mathcal{D}^{-1/2}\mathcal{R} \geq \frac{\beta}{1+\beta}\mathcal{R}$ assuming we can show $\min(\text{eig}(\mathcal{R}\mathcal{D}^{-1/2}\mathcal{B}\mathcal{D}^{-1/2}\mathcal{R})) \geq 0$. That is, we need to show that all those eigenvalues of $\mathcal{D}^{-1/2}\mathcal{B}\mathcal{D}^{-1/2}$ that are below 1 are also above $\beta/(1+\beta)$.

Without loss of generality, we normalize the common risk aversion to be $\alpha = 1$. In order to determine the eigenvalues of the matrix $\mathcal{D}^{-1/2}\mathcal{B}\mathcal{D}^{-1/2}$, we note that they coincide with the eigenvalues of $\mathcal{D}^{-1/2}(\mathcal{D}^{-1/2}\mathcal{B}\mathcal{D}^{-1/2})\mathcal{D}^{1/2} = \mathcal{D}^{-1}\mathcal{B} = \mathcal{E}^{\mathcal{M}}$.[33]

[33] For any invertible matrix $\mathcal{T}$ and any matrix $\mathcal{C}$, the eigenvalues of $\mathcal{T}^{-1}\mathcal{C}\mathcal{T}$ coincide with those of $\mathcal{C}$.

Fix a $\nu \in \mathbb{R}$ and let us calculate the inverse of $(\mathcal{E}^{\mathcal{M}} - \nu\mathrm{Id})$. Our goal is to show that it is invertible for all ν below $\beta/(1+\beta)$.

Writing equation $(\mathcal{E}^{\mathcal{M}} - \nu\mathrm{Id})X = Z$ with $X = (q_i^0)$ we get $((1-\nu)\mathrm{Id} - \bar{\Gamma}_i)\, q_i^0 + \bar{\mathcal{S}}_i \mathbf{Q}^* = Z_i$, implying that $q_i^0 = ((1-\nu)\mathrm{Id} - \bar{\Gamma}_i)^{-1}(Z_i - \bar{\mathcal{S}}_i \mathbf{Q}^*)$. Recalling the definition of $\mathbf{Q}^*$, we get

$$\mathbf{Q}^* = B^{-1}\sum_i \bar{\mathcal{S}}_i \alpha_i \mathbf{V} \bar{q}_i^0 = B^{-1}\sum_i \bar{\mathcal{S}}_i \alpha_i \mathbf{V}\big((1-\nu)\,\mathrm{Id} - \bar{\Gamma}_i\big)^{-1}(Z_i - \bar{\mathcal{S}}_i \mathbf{Q}^*),$$

which is equivalent to

$$\mathbf{Q}^* = \Big(B + \sum_i \bar{\mathcal{S}}_i \alpha_i \mathbf{V}\,((1-\nu)\mathrm{Id} - \bar{\Gamma}_i)^{-1}\bar{\mathcal{S}}_i\Big)^{-1} \sum_i \bar{\mathcal{S}}_i \alpha_i \mathbf{V}\,((1-\nu)\mathrm{Id} - \bar{\Gamma}_i)^{-1} Z_i,$$

and $(\mathcal{E}^{\mathcal{M}} - \nu\mathrm{Id})$ is invertible if and only if the right-hand side of this equation is well defined and finite.

The first key observation is that for each i, the matrix $\bar{\mathcal{S}}_i \alpha_i \mathbf{V}((1-\nu)\mathrm{Id} - \bar{\Gamma}_i)^{-1}\bar{\mathcal{S}}_i$ is symmetric. This follows because

$$\begin{aligned} \mathbf{V}_{N(i)}((1-\nu)\mathrm{Id} - \Gamma_i)^{-1} &= \mathbf{V}_{N(i)}((1-\nu)\mathrm{Id} - \mathcal{S}_i \mathbf{V}_{N(i)})^{-1} \\ &= ((1-\nu)\,\mathbf{V}_{N(i)}^{-1} - \mathcal{S}_i)^{-1}. \end{aligned} \tag{34}$$

The second key observation is that for $\nu \le \frac{\beta}{1+\beta}$, the matrix $\mathbf{V}_{N(i)}((1-\nu)\mathrm{Id} - \Gamma_i)^{-1}$ is positive semidefinite. This is equivalent to the claim that $(1-\nu)\mathbf{V}_{N(i)}^{-1} - \mathcal{S}_i \ge 0$. The latter follows because by Theorem 2, $\mathcal{S}_i \le \frac{1}{1+\beta}\mathbf{V}_{N(i)}^{-1}$. Thus, for $\nu < \frac{\beta}{1+\beta}$, the matrix $B + \sum_i \bar{\mathcal{S}}_i \alpha_i \mathbf{V}\,((1-\nu)\mathrm{Id} - \bar{\Gamma}_i)^{-1}\bar{\mathcal{S}}_i$ is positive definite and hence is invertible.

Fix now a market structure. Let $\mathcal{N}$ be the subset of exchanges for which the market structure is not equivalent to that of a centralized market (see Malamud and Rostek 2016 for a full characterization of such market structures). Then, on the complement of these exchanges, a straightforward application of the iteration procedure used for equilibrium construction in the proof of Theorem 1 implies that the price impacts on that subset of exchanges are strictly below β. Using continuity arguments it is then possible to show that the inequality $\mathcal{S}_i \le \frac{1}{1+\beta_i}\mathbf{V}_{N(i)}^{-1}$ still holds when the heterogeneity in risk aversion is sufficiently small, and then the arguments above imply the required welfare comparison result. The proof is complete. ∎

PROOF OF PROPOSITION 4:

Without loss of generality, we normalize the asset's variance to 1. Let us split the market in two so that agents $I-2, I-1, I$ trade in a separate exchange numbered 2. Furthermore, without loss of generality let us normalize $\alpha_{I-3} = 1$. Our goal is to show that for sufficiently small $\varepsilon \equiv \alpha_3$ there exists a vector of initial endowments such that the split market has a higher total welfare than the centralized market. For simplicity, we assume that $\alpha_1 = \alpha_2 = \alpha_3 = \varepsilon$. Denote by λ_i and λ_i' the

price impacts in the two market structures. Then, a direct calculation implies that $b = \frac{3}{2}\varepsilon^{-1} + \theta_0$, $b' = \frac{3}{2}\varepsilon^{-1} + \theta_0'$, and

$$\lambda_i = \frac{2\varepsilon}{3} + \frac{4}{9}\varepsilon^2\frac{1-\theta_0\alpha_i}{\alpha_i} + O(\varepsilon^3), \quad \lambda_i' = \frac{2\varepsilon}{3} + \frac{4}{9}\varepsilon^2\frac{1-\theta_0'\alpha_i}{\alpha_i} + O(\varepsilon^3), \quad i > 3,$$

while $\lambda_2 = \varepsilon - \theta_0\varepsilon^2\frac{4}{5}$, $\lambda_2' = \varepsilon - \theta_0'^2\frac{4}{5}$, where $\theta_0 = \frac{5}{2}\sum_{i>3}\alpha_i^{-1}$, $\theta_0' = \frac{5}{2}\sum_{3<i<I-2}\alpha_i^{-1}$. Denote by $s_i \equiv (\lambda_i + \alpha_i)^{-1}$ the lifted slope of agent i's demand. Suppose also that $q_i^0 = 0$ for $i = I-2, I-1, I$. Then, the equilibrium allocation is given by

$$q_j + q_j^0 = q_j^0 + s_j\Big(b^{-1}\sum_i s_i\alpha_i q_i^0 - \alpha_j\bar{q}_j^0\Big). \tag{35}$$

Let $\mathcal{D} = \mathrm{diag}(\alpha_i)$, $\mathcal{B} \equiv (s_i b^{-1} s_j)_{i,j}$, and $\mathcal{P} = \mathrm{diag}(s_i)$. Then, denoting $\mathcal{E} = \mathrm{Id} + (\mathcal{B}-\mathcal{P})\mathcal{D}$, we get $\mathcal{E}(q_j^0)_j = (q_j + q_j^0)_j$, and the equilibrium total welfare loss can be rewritten as $((q_i^0))^T\mathcal{A}(q_i^0)$, $\mathcal{A} = (\mathrm{Id} + \mathcal{D}(\mathcal{B}-\mathcal{P}))\mathcal{D}(\mathrm{Id} + (\mathcal{B}-\mathcal{P})\mathcal{D})$. Similarly, we use $\mathcal{A}'$, $\mathcal{B}'$, $\mathcal{P}'$ to denote the corresponding objects in the split market. Our goal is to show that there exists a vector of initial endowments $(q_i^0)_i$ such that $(q_i^0)_i^T\mathcal{A}(q_i^0)_i > (q_i^0)_i^T\mathcal{A}'(q_i^0)_i$. That is, the total welfare loss is higher in the centralized market. Equivalently, we need to show that $\max(\mathrm{eig}(\mathcal{A}-\mathcal{A}')) > 0$. Denote $\eta_i \equiv \alpha_i s_i$. Then,

$$\begin{aligned}\mathcal{A} &= \mathcal{D} + 2\mathcal{D}\mathcal{B}\mathcal{D} - 2\mathcal{D}\mathcal{P}\mathcal{D} + \mathcal{D}\mathcal{B}\mathcal{D}\mathcal{B}\mathcal{D} - \mathcal{D}\mathcal{B}\mathcal{D}\mathcal{P}\mathcal{D} - \mathcal{D}\mathcal{P}\mathcal{B}\mathcal{D}\mathcal{B} + \mathcal{D}\mathcal{P}\mathcal{D}\mathcal{P}\mathcal{D} \\ &= ((\alpha_i - 2\alpha_i\eta_i + \alpha_i\eta_i^2)\mathbf{1}_{i=j})_{i,j} + \Big(2b^{-1} + b^{-2}\sum_\ell s_\ell\eta_\ell\Big)(\eta_i\eta_j)_{i,j} \\ &\quad - b^{-1}(\eta_i\eta_j(\eta_i+\eta_j))_{i,j}.\end{aligned} \tag{36}$$

Hence, by direct (but tedious) calculation, the difference between $\mathcal{A}$ and $\mathcal{A}'$ (restricted to endowments corresponding to agents in the first exchange) is approximately given by

$$\begin{aligned}(\theta_0 - \theta_0')^{-1}(\mathcal{A}-\mathcal{A}') \approx\ & -\frac{\varepsilon^2}{5}(\mathbf{1}_{i=j}\mathbf{1}_{i\le 3})_{i,j} - \frac{11}{15}\varepsilon^2(\eta_i^*\eta_j^*)_{i,j} \\ & + \varepsilon^2((1-\mathbf{1}_{i>3,j>3})\kappa_{ij})_{i,j} + \frac{4}{9}\varepsilon^2(\eta_i^*\eta_j^*(\eta_i^*+\eta_j^*))_{i,j}\end{aligned} \tag{37}$$

for some $\kappa_{ij} \in \mathbb{R}$, where $\eta_i^* = 1 - 0.5\,\mathbf{1}_{i\le 3}$. Hence, the diagonal elements of this matrix for all $i > 3$ are positive, and therefore it cannot be negative semidefinite. ∎

PROOF OF PROPOSITION 5:

Part (i) is proved in the text. Part (ii): without loss of generality, we may assume that there are 5 agents in the market. We also assume for simplicity that three low risk aversion agents have the same risk aversion α_2, and the two high risk aversion agents have the same risk aversion $\alpha_1 > \alpha_2$. Consider the two market structures similar to those from Example 5. Namely, we assume that in the more centralized market, two of the less risk-averse agents trade only asset 2, while two high risk

aversion agents and the third low risk aversion agent trade both assets. Then, price impacts satisfy

$$(38)\quad \Lambda_1 = \left(\begin{pmatrix}0 & 0\\ 0 & 2(\Lambda_2+\alpha_2)^{-1}\end{pmatrix} + (\Lambda_1+\alpha_1\mathbf{V})^{-1} + (\Lambda_3+\alpha_2\mathbf{V})^{-1}\right)^{-1},$$

$$\Lambda_3 = \left(\begin{pmatrix}0 & 0\\ 0 & 2(\Lambda_2+\alpha_2)^{-1}\end{pmatrix} + 2(\Lambda_1+\alpha_1\mathbf{V})^{-1}\right)^{-1},$$

$$\Lambda_2 = \left(\left(\begin{pmatrix}0 & 0\\ 0 & (\Lambda_2+\alpha_2)^{-1}\end{pmatrix} + 2(\Lambda_1+\alpha_1\mathbf{V})^{-1} + (\Lambda_3+\alpha_2\mathbf{V})^{-1}\right)^{-1}\right)_{22}.$$

In the less centralized market, we prohibit the third low risk aversion agent from trading asset 2, so that

$$(39)\quad \Lambda_1' = \left(\begin{pmatrix}(\Lambda_3'+\alpha_2)^{-1} & 0\\ 0 & 2(\Lambda_2'+\alpha_2)^{-1}\end{pmatrix} + (\Lambda_1'+\alpha_1\mathbf{V})^{-1}\right)^{-1},$$

$$\Lambda_3' = \left(\left(\begin{pmatrix}0 & 0\\ 0 & 2(\Lambda_2'+\alpha_2)^{-1}\end{pmatrix} + 2(\Lambda_1'+\alpha_1\mathbf{V})^{-1}\right)^{-1}\right)_{11},$$

$$\Lambda_2' = \left(\left(\begin{pmatrix}(\Lambda_3'+\alpha_2)^{-1} & 0\\ 0 & (\Lambda_2'+\alpha_2)^{-1}\end{pmatrix} + 2(\Lambda_1'+\alpha_1\mathbf{V})^{-1}\right)^{-1}\right)_{22}.$$

When the correlation ρ is 0, the price impact of class 1 agents satisfies

$$\Lambda_* = \operatorname{diag}(\lambda_{11},\lambda_{21}),\quad \Lambda_*' = \operatorname{diag}(\lambda_{11}',\lambda_{21}'),$$

where $\lambda_{i1} = (2-\alpha_1 b_i+\sqrt{(\alpha_1 b_i)^2+4})/2b_i$, $\lambda_{i1}' = (2-\alpha_1 b_i'+\sqrt{(\alpha_1 b_i')^2+4})/2b_i'$, where

$$(40)\quad 2\left(2+\alpha_1 b_1+\sqrt{(\alpha_1 b_1)^2+4}\right)^{-1} + \left(2+\alpha_2 b_1+\sqrt{(\alpha_2 b_1)^2+4}\right)^{-1} = 1/2,$$

$$2\left(2+\alpha_1 b_2+\sqrt{(\alpha_1 b_2)^2+4}\right)^{-1} + 3\left(2+\alpha_2 b_2+\sqrt{(\alpha_2 b_2)^2+4}\right)^{-1} = 1/2.$$

At the same time, in the less centralized market, we have

$$(41)\quad 2\left(2+\alpha_1 b_1'+\sqrt{(\alpha_1 b_1')^2+4}\right)^{-1} + \left(2+\alpha_2 b_1'+\sqrt{(\alpha_2 b_1')^2+4}\right)^{-1} = 1/2,$$

$$2\left(2+\alpha_1 b_2'+\sqrt{(\alpha_1 b_2')^2+4}\right)^{-1} + 2\left(2+\alpha_2 b_2'+\sqrt{(\alpha_2 b_2')^2+4}\right)^{-1} = 1/2.$$

In particular, $b_1 = b_1'$ and hence liquidity for the first asset is the same in both markets. At the same time, $b_2' < b_2$ and hence liquidity for the second asset is higher in the more centralized market. It is possible to show that when $\alpha_2 \to 0$, $b_1 = \frac{1}{\alpha_1}\frac{8}{3} + O(\alpha_2)$ and $\lambda_{11} = \lambda_{11}' = \frac{\alpha_1}{3} + O(\alpha_2)$ while $b_2 \approx \frac{3}{4\alpha_2}$

and $\lambda_{21} \approx \frac{2\alpha_2}{3}$. At the same time, $b_2' \approx \frac{2^{3/2}}{\sqrt{\alpha_1\alpha_2}}$ and $\lambda_{21}' \approx \frac{\sqrt{\alpha_1\alpha_2}}{2^{3/2}}$. Suppose now that $\rho \approx \rho^*\sqrt{\alpha_2}$. Then, $\Gamma_1 = \hat{\Gamma} + \sqrt{\alpha_2}\Gamma_* + O(\alpha_2)$, $\Gamma_1' \approx \hat{\Gamma} + \sqrt{\alpha_2}\Gamma_*' + O(\alpha_2)$, where $\hat{\Gamma} = \hat{S} - 0.5\alpha_1\hat{S}^2$ with $\hat{S} = \operatorname{diag}((\frac{\alpha_1}{3} + \alpha_1)^{-1}, \alpha_1^{-1})$. Our goal is to show that $\Gamma_* \not\geq \Gamma_*'$ for some values of ρ^*.

The first observation is that by Proposition 7 in the online Appendix, to the highest order (i.e., $\sqrt{\alpha_2}$), Λ_1 only depends on ρ through $1 - \rho^2 = 1 + O(\alpha_2)$ and hence $\Lambda = \Lambda^* + O(\alpha_2)$. At the same time, ignoring the terms of the order of α_2, we have that

$$\Lambda_1' = \left(\begin{pmatrix}(\Lambda_3')^{-1} & 0\\ 0 & 2(\Lambda_2')^{-1}\end{pmatrix} + (\Lambda_1' + \alpha_1\mathbf{V})^{-1}\right)^{-1},$$

$$\Lambda_3' = \left(\left(\begin{pmatrix}0 & 0\\ 0 & 2(\Lambda_2')^{-1}\end{pmatrix} + 2(\Lambda_1' + \alpha_1\mathbf{V})^{-1}\right)^{-1}\right)_{11},$$

$$\Lambda_2' = \left(\left(\begin{pmatrix}(\Lambda_3')^{-1} & 0\\ 0 & (\Lambda_2')^{-1}\end{pmatrix} + 2(\Lambda_1' + \alpha_1\mathbf{V})^{-1}\right)^{-1}\right)_{22}.$$

Substituting $\mathbf{V} = \mathrm{Id} + \rho^*\sqrt{\alpha_2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$ and using the implicit function theorem, we get $\Lambda_1' \approx \Lambda_* + \sqrt{\alpha_2}\tilde{\Lambda}$. Calculating Γ_*' using this expression yields the required result.

To prove the last claims of Proposition 5, we will need the following auxiliary result.

LEMMA 4: *Within Proposition* 7, *let the equilibrium price impact of trader* 1 *be* $\tilde{\Lambda}_1$. *Then, any* $\hat{\Lambda}_1 \leq \Lambda_1$ *can be attained as an equilibrium price impact by adding an additional trader to the exchange.*

PROOF:

See the online Appendix.

Lemma 4 implies that by adding/removing traders to/from the illiquid exchange, the required changes in price impact can be achieved. The claim now follows from Proposition 2(ii).[34] Assume that the endowments are such that there is no inter-class trade and also no intra-class trade, except for agents of class 1. As we discuss in the main text, this can be done independently of price impacts. For simplicity, we may assume that the total endowment of each class is 0, so that $Q = 0$. We can also assume that class 1 consists of 2 traders with endowments q_1^0 and $-q_1^0$, so that we are free to choose this endowment without affecting aggregate risk. ∎

[34] While Proposition 2(ii) only claims that Υ may be nonmonotone in Λ, it is straightforward to show that it is in fact always nonmonotone in Λ when Λ is not proportional to $\mathbf{V}_{N(i)}$.

REFERENCES

Afonso, Gara, Anna Kovner, and Antoinette Schoar. 2013. "Trading Partners in the Interbank Lending Market." Federal Reserve Bank of New York Staff Report 620.

Afonso, Gara, and Ricardo Lagos. 2012. "Trade Dynamics in the Market for Federal Funds." Federal Reserve Bank of New York Staff Report 549.

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (3): 488–500.

Anderson, William N., Jr. 1971. "Shorted Operators." *SIAM Journal of Applied Mathematics* 20 (3): 520–25.

Anderson, William N., Jr., and R. J. Duffin. 1969. "Series and Parallel Addition of Matrices." *Journal of Mathematical Analysis and Applications* 26 (3): 576–94.

Angel, James J., Lawrence E. Harris, and Chester S. Spatt. 2011. "Equity Trading in the 21st Century." *Quarterly Journal of Finance* 1 (1): 1–53.

Atkeson, Andrew G., Andrea L. Eisfeldt, and Pierre-Olivier Weill. 2015. "Entry and Exit in OTC Derivatives Markets." *Econometrica* 83 (6): 2231–92.

Babus, Ana, and Péter Kondor. 2016. "Trading and Information Diffusion in Over-the-Counter Markets." Unpublished.

Bech, Morten L., and Enghin Atalay. 2010. "The Topology of the Federal Funds Market." *Physica A: Statistical Mechanics and Its Applications* 389 (22): 5223–46.

Berge, Claude. 1973. *Graphs and Hypergraphs.* Amsterdam: North-Holland.

Biais, Bruno, Christophe Bisière, and Chester Spatt. 2010. "Imperfect Competition in Financial Markets: An Empirical Study of Island and Nasdaq." *Management Science* 56 (12): 2237–50.

Biais, Bruno, and Richard C. Green. 2007. "The Microstructure of the Bond Market in the 20th Century." IDEI Working Paper 482.

Blume, Lawrence E., David Easley, Jon Kleinberg, and Éva Tardos. 2009. "Trading Networks with Price-Setting Agents." *Games and Economic Behavior* 67 (1): 36–50.

Bramoullé, Yann, Rachel Kranton, and Martin D'Amours. 2014. "Strategic Interaction and Networks." *American Economic Review* 104 (3): 898–930.

Choi, Syngjoo, Andrea Galeotti, and Sanjeev Goyal. 2017. "Trading in Networks: Theory and Experiment." *Journal of the European Economic Association* 15 (4): 784–817.

Cocco, João F., Francisco J. Gomes, and Nuno C. Martins. 2009. "Lending Relationships in the Interbank Market." *Journal of Financial Intermediation* 18 (1): 24–48.

Condorelli, Daniele, Andrea Galeotti, and Ludovic Renou. 2017. "Bilateral Trading in Networks." *Review of Economic Studies* 84 (1): 82–105.

Corominas-Bosch, Margarida. 2004. "Bargaining in a Network of Buyers and Sellers." *Journal of Economic Theory* 115 (1): 35–77.

Craig, Ben, and Peter von Goetz. 2014. "Interbank Tiering and Money Center Banks." *Journal of Financial Intermediation* 23 (3): 322–47.

Donoghue, William F., Jr. 1974. *Monotone Matrix Functions and Analytic Continuation*. New York: Springer-Verlag.

Duffie, Darrell. 2012. *Dark Markets: Asset Pricing and Information Transmission in Over-the-Counter Markets.* Princeton, NJ: Princeton University Press.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen. 2005. "Over-the-Counter Markets." *Econometrica* 73 (6): 1815–47.

Duffie, Darrell, Semyon Malamud, and Gustavo Manso. 2009. "Information Percolation with Equilibrium Search Dynamics." *Econometrica* 77 (5): 1513–74.

Duffie, Darrell, Semyon Malamud, and Gustavo Manso. 2014. "Information Percolation in Segmented Markets." *Journal of Economic Theory* 153: 1–32.

Elliott, Matthew. 2011. "Search with Multilateral Bargaining." Unpublished.

Elliott, Matthew. 2015. "Inefficiencies in Networked Markets." *American Economic Journal: Microeconomics* 7 (4): 43–82.

Gale, Douglas. 1986a. "Bargaining and Competition Part I: Characterization." *Econometrica* 54 (4): 785–806.

Gale, Douglas. 1986b. "Bargaining and Competition Part II: Existence." *Econometrica* 54 (4): 807–818.

Gale, Douglas, and Shachar Kariv. 2007. "Financial Networks." *American Economic Review* 97 (2): 99–103.

Gofman, Michael. 2011. "A Network-Based Analysis of Over-the-Counter Markets." Unpublished.

Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill. 2014. "Heterogeneity in Decentralized Asset Markets." Unpublished.

Hugonnier, Julien, S. Malamud, and E. Trubowitz. 2012. "Endogenous Completeness of Diffusion Driven Equilibrium Markets." *Econometrica* 80 (3): 1249–70.

Jackson, Matthew O., and Yves Zenou. 2014. "Games on Networks." In *Handbook of Game Theory,* Vol. 4, edited by H. Peyton Young and Shmuel Zamir, 95–164. New York: Elsevier Science.

Knight Capital Group, eds. 2010. *Current Perspectives on Modern Equity Markets: A Collection of Essays by Financial Industry Experts*. Washington, DC: Knight Capital Group.

Kranton, Rachel E., and Deborah F. Minehart. 2001. "A Theory of Buyer-Seller Networks." *American Economic Review* 91 (3): 485–508.

Kyle, Albert S. 1989. "Informed Speculation and Imperfect Competition." *Review of Economic Studies* 56 (3): 317–55.

Lagos, Ricardo, and Guillaume Rocheteau. 2009. "Liquidity in Asset Markets with Search Frictions." *Econometrica* 77 (2): 403–26.

Lagos, Ricardo, Guillaume Rocheteau, and Pierre-Olivier Weill. 2011. "Crises and Liquidity in Over-the-Counter Markets." *Journal of Economic Theory* 146 (6): 2169–2205.

Lambert, Nicholas, Michael Ostrovsky, and Mikhail Panov. 2016. "Strategic Trading in Informationally Complex Environments." Unpublished.

Li, Dan, and Norman Schürhoff. 2012. "Dealer Networks." Unpublished.

Malamud, Semyon, and Marzena Rostek. 2016. "Noncompetitive Decentralized Markets." Unpublished.

Manea, Mihai. 2011. "Bargaining in Stationary Networks." *American Economics Review* 101 (5): 2042–80.

Nava, Francesco. 2015. "Efficiency in Decentralized Oligopolistic Markets." *Journal of Economic Theory* 157: 315–48.

O'Hara, Maureen, and Mao Ye. 2011. "Is Market Fragmentation Harming Market Quality?" *Journal of Financial Economics* 100 (3): 459–74.

Peivandi, Ahmad, and Rakesh V. Vohra. 2013. "On Fragmentation of Markets." Unpublished.

Penrose, R. 1955. "A Generalized Inverse for Matrices." *Mathematical Proceedings of the Cambridge Philosophical Society* 51 (3): 406–13.

Rahi, Rohit, and Jean-Pierre Zigrand. 2013. "Market Quality and Contagion in Fragmented Markets." Unpublished.

Rostek, Marzena, and Marek Weretka. 2015a. "Dynamic Thin Markets." *Review of Financial Studies* 28 (10): 2946–92.

Rostek, Marzena, and Marek Weretka. 2015b. "Information and Strategic Behavior." *Journal of Economic Theory* 158 (PB): 536–57.

Schapiro, Mary L. 2010. "Strengthening Our Equity Market Structure." Speech at Economic Club of New York, September 7. http://www.sec.gov/news/speech/2010/spch090710mls.htm.

Vayanos, Dimitri. 1999. "Strategic Trading and Welfare in a Dynamic Market." *Review of Economic Studies* 66 (2): 219–54.

Vayanos, Dimitri, and Pierre-Olivier Weill. 2008. "A Search-Based Theory of the On-the-Run Phenomenon." *Journal of Finance* 63 (3): 1361–98.

Vives, Xavier. 2011. "Strategic Supply Function Competition with Private Information." *Econometrica* 79 (6): 1919–66.

Weill, Pierre-Olivier. 2008. "Liquidity Premia in Dynamic Bargaining Markets." *Journal of Economic Theory* 140 (1): 66–96.

Weretka, Marek. 2011. "Endogenous Market Power." *Journal of Economic Theory* 146 (6): 2281–2306.