Benign Granularity in Asset Markets*

Sergei Glebkin, Semyon Malamud, and Alberto Teguia

March 3, 2025

Abstract

We develop a tractable model to study how asset concentration among a few large investors impacts asset prices and liquidity. Consistent with existing empirical evidence: (i) greater concentration is associated with higher volatility and returns, and (ii) large investors' turnover share is smaller than their proportion of total wealth. Surprisingly, higher concentration enhances liquidity, aligning with our new empirical findings. We show that increased concentration can benefit all investors in sufficiently non-competitive markets. We link the wedge between competitive and non-competitive outcomes to the Herfindahl-Hirschman Index measuring wealth concentration. The wedge can remain positive even in large markets.

JEL Classification: D21, G31, G32, G35, L11

Keywords: Market Liquidity, Funding Liquidity, Price Impact, Strategic Trading, Wealth Effects

*Sergei Glebkin (glebkin@insead.edu) is at INSEAD, Semyon Malamud (semyon.malamud@epfl.ch) is at EPFL, and Alberto Teguia (alberto.mokakteguia@sauder.ubc.ca) is at UBC Sauder.

1 Introduction

Modern markets are increasingly dominated by large institutional investors.[1] Moreover, a few firms, such as BlackRock, Vanguard, and Fidelity in the US, have a disproportionate share of the total assets under management (AUM).[2] The concentration of institutional ownership has raised regulatory concerns that trades of large institutions, due to their significant price impact, could lead to excess volatility or even pose systemic risks through spillover effects on the values of portfolios held by other investors. Empirical studies offer partial support for these concerns and show that changes in concentration are indeed linked to changes in volatility (Ben-David et al., 2021) and asset returns (Massa, Schumacher, and Wang, 2021).

In this paper, we propose an equilibrium model to study how changes in institutional ownership concentration affect asset markets. We analyze the implications for asset returns, volatility, and liquidity. We also examine the welfare consequences of changes in concentration. We uncover a surprising and benign aspect of concentration: increased concentration is associated with higher liquidity. We provide suggestive empirical evidence supporting this prediction. Furthermore, we show that this enhanced liquidity can help translate increases in concentration into improved welfare, driven by better intertemporal consumption smoothing and improved risk-sharing among traders.

Our model incorporates three main features: (i) *wealth effects*, allowing changes in funds' AUM to influence equilibrium outcomes; (ii) *wealth heterogeneity*, providing generality and flexibility in the AUM distribution across funds; and (iii) *non-competitive traders*, enabling changes in concentration to have aggregate effects and capturing the behavior of large traders who not only have price impact but also account for it.[3]

[1]Since 1980, the share of U.S. stocks owned by the top 10 institutions has more than quadrupled, reaching 26.5% by 2016 (Ben-David, Franzoni, Moussawi, and Sedunov, 2021).

[2]For instance, Vanguard and Fidelity together accounted for 30% of the mutual fund industry's market share in 2018 (Tjornehoj, 2018).

[3]Many institutional investors, such as J.P. Morgan and Citigroup, maintain in-house "optimal execution" desks to devise trading strategies that minimize costs. Other investors rely on software and services offered by specialized trading firms.

In particular, there is a single trading period where two groups of traders, *liquidity providers (LPs)* and *liquidity demanders (LDs)*, trade multiple risky assets with arbitrarily distributed asset payoffs. LPs share identical Epstein-Zin preferences with a unit elasticity of intertemporal substitution (EIS), enabling wealth effects while maintaining analytical tractability. LPs are symmetrically informed, and the absence of information asymmetry implies that price impact arises solely from inventory risk.[4] Linking to our empirical motivation, we interpret LPs as funds and their initial wealth as the size of the funds' AUM. Heterogeneity in LPs' initial wealth allows us to examine how changes in AUM concentration across traders affect equilibrium outcomes. Trading is structured as a uniform-price double auction, where LPs submit demand functions specifying their desired asset quantities as functions of asset prices. All trades are executed at prices that clear the market. When choosing their optimal demand schedules, LPs account for their price impact. LPs compete to provide liquidity to LDs, who submit market orders. In the first part of the paper, we remain agnostic about LDs' preferences, expressing equilibrium quantities as functions of aggregate liquidity demand. For welfare analysis, we assume LDs have the same unit-elastic Epstein-Zin preferences as LPs.

We begin by examining the competitive benchmark, where LPs take prices as given. In this setting, classic aggregation results imply that the distribution of AUM across LPs has *no effect* on key quantities of interest: returns, volatility, and liquidity. Consequently, addressing the regulatory concerns and empirical findings discussed earlier requires a model that incorporates non-competitive trading.

To this end, we analyze the non-competitive equilibrium. Our equilibrium features a tractable *scale-symmetric* structure, where the demand schedules of different traders are proportional to a common schedule, with scaling constants determined by the AUM distribution across LPs. The cross-section of demand schedules exhibits several notable properties.

First, in equilibrium, larger investors engage in larger trades and provide more liquidity. This is intuitive: with larger balance sheets, they have greater trading needs and a higher

[4]We abstract from asymmetric information for tractability.

capacity to provide liquidity. Second, consistent with Koijen and Yogo (2019), larger investors also experience greater price impact. This arises because their price impact depends on the liquidity provided by other investors, who, on average, provide less of it. Finally, unlike in the competitive model—where an investor's turnover share matches their wealth share—in the non-competitive setting, the largest investors have turnover shares smaller than their wealth shares, while the smallest investors exhibit the opposite pattern. This, too, aligns with the findings of Koijen and Yogo (2019) and is explained in our model by the larger price impact faced by bigger investors, prompting them to adjust their demands more significantly away from the competitive benchmark—where turnover shares align with wealth shares—compared to smaller investors.

Comparing equilibrium returns and volatility in competitive and non-competitive settings, we find that both returns and volatility are higher in the non-competitive equilibrium. This is intuitive: when LPs exercise market power, they tilt returns (both realized and expected) in their favor, scaling them up. This scale-up naturally leads to greater volatility.

For the same reason, in a non-competitive equilibrium, an increase in AUM concentration—whether due to a merger of funds or fund flows from smaller to larger funds—amplifies the market power of large traders, resulting in higher returns and volatility. These findings align with the empirical evidence in Massa et al. (2021) (for returns) and Ben-David et al. (2021) (for volatility).

A surprising result of our theory is that liquidity (defined as a more price-elastic aggregate demand) is higher in the non-competitive equilibrium compared to the competitive one or when AUM concentration among LPs increases. The intuition behind this counter-intuitive result is non-trivial and is unique to a model with wealth effects, as we now explain. For simplicity, we focus our discussion on a single asset case, but our model allows for many assets with arbitrarily distributed payoffs.

The standard optimality condition for the non-competitive demand (Ausubel, Cramton, Pycia, Rostek, and Weretka, 2014) implies the following decomposition

$$\text{non-competitive inverse demand}(q) = \text{competitive inverse demand}(q) - \text{price impact}(q) \cdot q.$$

The inverse demand for q units of the asset can be thus decomposed into the competitive demand and the *demand reduction* component, price impact$(q) \cdot q$. In traditional linear models, where the price impact is constant, the demand reduction increases with q. Consequently, the slope of the inverse demand becomes larger (in absolute value) due to the demand reduction component. This implies that LPs' demands become less elastic (since demand elasticity is inversely related to the slope of the inverse demand). As a result, traders provide less liquidity, leading to an overall decline in market liquidity. In contrast, in our model, the demand reduction *decreases* with q. Consequently, the slope of the inverse demand, *becomes smaller* due to the demand reduction component. This implies that demands become *more elastic*, leading to an overall *increase* in market liquidity.[5]

In the empirical part of the paper, we examine the relationship between institutional ownership concentration—measured by the Herfindahl-Hirschman Index (HHI) of fund AUM distribution—and key aggregate quantities: market volatility (proxied by the VIX) and illiquidity (proxied by Amihud's lambda). Consistent with previous empirical findings (Ben-David et al., 2021) and in line with our theory, we find a positive *predictive* relationship between VIX and past changes in HHI. Further, in line with our theory and the discussion above, we find that aggregate Amihud's lambda is positively associated with past changes in HHI.

We then proceed with a welfare analysis, focusing on the effects of a merger between two LPs. We demonstrate that when the asset market is sufficiently non-competitive, *all traders* could *benefit* from increased concentration.[6] LDs benefit from the merger due to improved

[5]The property of demand reduction increasing in q is quite specific to linear models and does not hold in general. In a model with symmetric CARA LPs and arbitrary payoff distributions, Glebkin, Malamud, and Teguia (2023a) demonstrate that demand reduction decreases with q for sufficiently large q. Glebkin, Malamud, and Teguia (2023b) further extend this result to accommodate general preferences.

[6]Glebkin and Kuong (2023) also demonstrate theoretically that mergers between large investors can benefit all traders. However, unlike our model and empirical evidence, their framework predicts that such mergers

liquidity, which enhances risk-sharing and consumption smoothing. Other LPs additionally benefit from more favorable prices, as increased concentration tilts prices in their favor.

We also explore whether markets become perfectly competitive as the number of traders grows indefinitely—a classic question in finance theory (see, e.g., Lee and Kyle (2018) and references therein). We show that the wedge between aggregate outcomes in competitive and non-competitive economies can be well approximated by the Herfindahl-Hirschman Index (HHI) of the AUM distribution across LPs. If HHI is zero in a large economy, the wedge disappears, implying perfect competition. However, if HHI remains positive, so does the wedge, meaning market power persists even in large economies. Our findings suggest that large economies are not necessarily competitive and provide a theoretical foundation for HHI, a key metric used by regulators to assess mergers.

The rest of the paper is organized as follows. Section 2 introduces the model. Section 3 examines the competitive benchmark, while Section 4 analyzes the non-competitive equilibrium. Section 5 studies the large-economy limit, and Section 6 explores welfare implications. Section 7 presents empirical evidence, followed by a review of the related literature in Section 8. Finally, Section 9 concludes.

2 The Model

There are two time periods $t \in \{0, 1\}$. A number $L > 1$ of strategic liquidity providers (LPs) trade assets with liquidity demanders (LDs) at $t = 0$. LPs are indexed by $i \in \{1, 2, \dots, L\}$. There are $N + 1$ assets, indexed by $k \in \{0, 1, 2, ..N\}$. An asset k is a claim to a terminal dividend δ_k, where asset 0 is a risk-free asset with $\delta_0 = 1$. All payoffs are collected in the vector $\delta = (\delta_k)_k$.[7]

harm liquidity. Moreover, their mechanism relies on private values and asymmetric information, which are not necessary for our result.

[7] All vectors are assumed to be column vectors unless stated otherwise.

LPs consume both at time 0 and time 1. Each LP i is endowed with $w_0^i = \alpha_i w_0$ units of the consumption good at time 0, where w_0 represents the total wealth of all LPs, and α_i denotes LP i's share of the total wealth. By definition, the shares sum to 1, i.e., $\sum_i \alpha_i = 1$. Suppose that investor i trades q_k units of asset k at time 0 at price P_k. Denote $q = (q_k)_k$ and $P = (P_k)_k$. LP i's time-0 consumption is then $c_0^i = \alpha_i w_0 - q^\top P$, and their time-1 consumption is $c_1^i = \delta^\top q$. LPs have Epstein-Zin (1989) preferences with the elasticity of intertemporal substitution (EIS) equal to 1 and the relative risk aversion parameter (RRA) equal to γ:

$$U_i(c_0^i, c_1^i) = \log(c_0^i) + \log E\left[\left(c_1^i\right)^{1-\gamma}\right]^{1/(1-\gamma)}.$$

We assume $\gamma \in (0, 1]$.[8] In the equation above and everywhere in the sequel, $E[\cdot]$ denotes expectation taken with respect to the distribution of δ.

In the first part of the paper, we abstract from LDs' preferences and assume that their aggregate trade is represented by an exogenous supply shock $Q \in \mathbb{R}^{N+1}$, that is uncertain to LPs and is independent of δ.[9] Later, when analyzing welfare, we endogenize Q by formulating the LDs' optimization problem. We impose the following technical restrictions on the distributions of δ and Q.

Assumption 1. *For every $y \in \operatorname{supp}\{Q\}$, $E\left[\left(\delta^\top y\right)^{-\gamma} \delta\right]$ is well defined: (i) $\delta^\top y > 0$, and (ii) $E\left[\left(\delta^\top y\right)^{-\gamma} \delta\right] < \infty$. Additionally, $\operatorname{Var}[\delta_k] < \infty$ for all k. We restrict the admissible portfolios q to be of the form $q = ty$, for some $t \in \mathbb{R}_+$ and $y \in \operatorname{supp}\{Q\}$. That is, q belong to the cone generated by Q.*

The assumption above ensures that traders' optimization problems and the equilibrium outcomes are well-defined. When Q is endogenized, condition (i) will be satisfied naturally.

[8]When $\gamma > 1$, the utility U_i becomes *convex* in c_1^i, which significantly complicates verifying the second-order conditions for optimality of the demands. Additionally, when $\gamma > 1$, the demand schedules for LPs become upward-sloping, which is counterfactual.

[9]Independence of δ and Q implies that LDs are uninformed. Uncertainty about Q is needed to rule out the multiplicity of equilibria (cf. Klemperer and Meyer (1989) and Vayanos (1999)). As in Klemperer and Meyer, our assumptions imply that equilibrium quantities will depend on the realization of Q but not its distribution.

Trading is organized as a uniform-price double auction: each LP i, $i \in \{1, 2, \dots L\}$ submits a demand function $D_i(P) : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ specifying the number of units of the assets they want to buy as a function of the prices of all assets. All trades are executed at prices P^* that clear the market, i.e., at a vector P^* such that $\sum_i D_i(P^*) = Q$. We examine Nash equilibria in demand schedules, where all LPs trade strategically, rationally anticipating the impact of their demand schedules on the market-clearing price.

3 Competitive benchmark

We start by characterizing the benchmark equilibrium where LPs take prices as given. LP i solves the following optimization problem:

$$\sup_q \left\{ \log(\alpha_i w_0 - q^\top P) + \log\left(E\left[\left(q^\top \delta\right)^{1-\gamma} \right]^{\frac{1}{1-\gamma}} \right) \right\}.$$

The first-order (necessary and sufficient) condition can be written as

$$\left(\alpha_i w_0 - q^\top P\right) \frac{E\left[\left(q^\top \delta\right)^{-\gamma} \delta \right]}{E\left[\left(q^\top \delta\right)^{1-\gamma} \right]} = P. \tag{1}$$

Pre-multiplying (1) by $q^\top$, we obtain $q^\top P = \alpha_i w_0 / 2$. Substituting this back into (1) yields a closed-form expression for LP i's inverse demand $I_i(q)$, specifying the prices they bid for a quantity vector q:

$$I_i(q) = \frac{\alpha_i w_0}{2} \frac{E\left[\left(q^\top \delta\right)^{-\gamma} \delta \right]}{E\left[\left(q^\top \delta\right)^{1-\gamma} \right]}.$$

We note several key properties of the competitive inverse demand:

1. $I_i(q)$ exhibit *scale symmetry*. This means that there exist scaling constants β_i, $i = \{1, 2, \dots, L\}$, and a function $I(q)$ such that $I_i(q) = \beta_i I(q)$ for all i. In our case, $\beta_i = \alpha_i$ and $I(q) = 0.5 w_0 \, E\left[\left(q^\top \delta\right)^{-\gamma} \delta \right] / E\left[\left(q^\top \delta\right)^{1-\gamma} \right]$.

2. The inverse demands are *homogeneous* in q. This means that there exists a constant k such that for any scalar $t \neq 0$ and any q, $I_i(tq) = t^k I_i(q)$. In our case, $k = -1$.

3. The inverse demands are *monotone* (strictly decreasing) functions of q. This means that $(I_i(q) - I_i(\hat{q}))^\top (q - \hat{q}) < 0$ for all $\hat{q} \neq q$.[10] Moreover, $I_i(q)$ is continuously differentiable, and the Jacobian $\nabla I_i(q)$ is non-degenerate.

We now turn to characterizing equilibrium allocations, prices, returns, volatility, and liquidity. Property 3 implies that the demand function, $D_i(P)$, which is the inverse of $I_i(q)$, is well-defined. Additionally, Properties 1 and 2 imply that the demand also exhibits scale symmetry and can be written as $D_i(P) = \alpha_i D(P)$. Given that supply is Q, the equilibrium allocation for trader i is $q_i = \alpha_i Q$. The equilibrium price is determined by $P(Q) = I_i(q_i) = I_i(\alpha_i Q)$, yielding the expression

$$P^c(Q) = \frac{w_0}{2} \frac{E\left[\left(Q^\top \delta\right)^{-\gamma} \delta\right]}{E\left[(Q^\top \delta)^{1-\gamma}\right]} \tag{2}$$

for the competitive equilibrium price $P^c(Q)$. Everywhere in the sequel, we use the superscript c to indicate the equilibrium outcomes in the competitive benchmark. The expected return on asset k is given by

$$\mu_k^c \equiv \frac{E[\delta_k]}{P_k^c(Q)} = \frac{2E[\delta_k] E\left[\left(Q^\top \delta\right)^{1-\gamma}\right]}{w_0 E\left[(Q^\top \delta)^{-\gamma} \delta_k\right]}. \tag{3}$$

The return volatility is given by

$$\sigma_k^c \equiv \mathrm{Var}[\delta_k / P_k^c(Q)]^{1/2} = \frac{2\mathrm{Var}[\delta_k]^{1/2} E\left[\left(Q^\top \delta\right)^{1-\gamma}\right]}{w_0 E\left[(Q^\top \delta)^{-\gamma} \delta_k\right]}. \tag{4}$$

Our measure of illiquidity is the sensitivity of equilibrium prices to supply shocks, a standard measure in the literature (see Vayanos and Wang (2012)). When there are multiple assets, the illiquidity is characterized by a matrix Λ whose (k, l)-th element measures the marginal effect

[10] We prove that $I_i(q)$ is monotone in Lemma 1 in the Appendix.

of a supply shock in asset l on the price of asset k,

$$\Lambda_{kl} = -\frac{\partial P_k}{\partial Q_l}.$$

Differentiating (2), we obtain

$$\Lambda^c(Q) = -\nabla P^c(Q) = \frac{w_0}{2}\left(\gamma \frac{E\left[\left(Q^\top \delta\right)^{-\gamma-1} \delta\delta^\top\right]}{E\left[\left(Q^\top \delta\right)^{1-\gamma}\right]} + (1-\gamma) P^c(Q) P^c(Q)^\top\right). \tag{5}$$

We summarize the properties of the unique competitive equilibrium in the following proposition.

Proposition 1. *The equilibrium prices, expected returns, return volatility, and illiquidity in the competitive case are given by equations* (2)*,* (3)*,* (4) *and* (5)*. These quantities are invariant to changes in the wealth distribution* $\{\alpha_i, i \in \{1, 2, \ldots, L\}\}$*.*

In our model, agents have proportional endowments and homothetic preferences. Classic aggregation results (see, e.g., Varian (1992)) imply that the economy features a representative agent and, hence, wealth distribution does not affect equilibrium prices. As we now show, this invariance breaks down when LPs act strategically.

4 Non-competitive equilibrium

In this section, we derive an equilibrium where LPs act strategically. Following the classical approach introduced by Wilson (1979), we model their strategic interactions as competition in demand schedules, $D_i(P)$. We adopt a guess-and-verify approach. We hypothesize that the strategic demands exhibit the three key properties discussed in the previous section: scale symmetry, homogeneity, and monotonicity. From this point forward, we refer to the Nash equilibrium in demand schedules that satisfy these properties simply as the equilibrium. Therefore, our objective is to identify a Nash equilibrium where LPs' demands satisfy:

$$D_i(P) = \beta_i D(P) \quad \text{for all } i,$$

where $D(P)$ is a strictly decreasing, continuously differentiable, homogeneous function, and $\beta_i > 0$ are constants. Without loss of generality, we normalize $\sum_i \beta_i = 1$.

Definition 1. *A tuple $(D(P), \beta)$, consisting of a function $D(P)\colon \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ and a vector $\beta \in \mathbb{R}^L_+$ with $\sum_i \beta_i = 1$, is an equilibrium if the following conditions hold:*

- *For any $i = 1, 2, \ldots, L$, if all other traders $j \neq i$ submit demands $D_j(P) = \beta_j D(P)$, then it is optimal for trader i to submit the demand $D_i(P) = \beta_i D(P)$.*
- *The function $D(P)$ is strictly decreasing, meaning $(D(P) - D(\hat{P}))^\top (P - \hat{P}) < 0$ for all $P \neq \hat{P}$. Additionally, $D(P)$ is continuously differentiable and has a non-degenerate Jacobian.*

Denoting the inverse of $D(P)$ by $I(q)$ (so that $I(D(P)) = P$), we can reformulate our ansatz in terms of inverse demands as follows:

$$I_i(q) = I(q/\beta_i) \quad \text{for all } i. \tag{6}$$

By Definition 1, we seek an equilibrium where $I(q)$ is a monotone, continuously differentiable, and homogeneous function of q with a non-degenerate Jacobian. An important insight from the literature on supply function competition (see, e.g., Kyle (1989) and Klemperer and Meyer (1989)) is that the equilibrium can be reformulated in terms of inverse residual demand curves faced by each trader. For a trader i, we denote the inverse residual demand by $P_i(q_i)$, which gives the vector of prices when agent i trades the quantity vector q_i. In this reformulation, the agent's objective is expressed as:

$$\sup_{q} \left\{ \log(\alpha_i w_0 - q^\top P_i(q)) + \log\left(E\left[\left(q^\top \delta\right)^{1-\gamma} \right]^{\frac{1}{1-\gamma}} \right) \right\}. \tag{7}$$

Importantly, the functions $P_i(q)$ are agent-specific, reflecting the heterogeneity assumed in the model.

4.1 Derivation of equilibrium

Recall the first-order condition (1) for a price-taking LP i trading a quantity vector q_i at prices $I_i(q_i)$:

$$I_i(q_i) = \left(\alpha_i w_0 - q_i^\top I_i(q_i)\right) \frac{E\left[\left(q_i^\top \delta\right)^{-\gamma} \delta\right]}{E\left[\left(q_i^\top \delta\right)^{1-\gamma}\right]}. \tag{8}$$

A strategic trader, however, accounts for the fact that they can influence prices. An important insight is that the solution to the problem (7) can be reformulated entirely in terms of the *price impact matrix*, $(\Lambda_i(q_i))_{kl} = \partial(P_i)_k / \partial(q_i)_l$, where the k,l-th element represents the sensitivity of the price of asset k to changes in LP i's demand for asset l. In vector notation, $\Lambda_i(q_i) = \nabla P_i(q_i) \in \mathbb{R}^{(N+1)\times(N+1)}$. With this definition, the first-order condition for (7) can be expressed as:

$$I_i(q_i) + \Lambda_i(q_i) q_i = \left(\alpha_i w_0 - q_i^\top I_i(q_i)\right) \frac{E\left[\left(q_i^\top \delta\right)^{-\gamma} \delta\right]}{E\left[\left(q_i^\top \delta\right)^{1-\gamma}\right]}. \tag{9}$$

To derive the equilibrium price impact, suppose LP i modifies her demand while other demands remain fixed at equilibrium. Thus, for other traders $j \neq i$, we have $D_j(P) = \beta_j D(p)$ and $I_j(q) = I(q/\beta_j)$. Suppose that total supply is Q. When trader i receives the quantity q_i, market clearing dictates that the remaining quantity $Q - q_i$ is allocated among the other traders proportionally to their scaling constants β_j. Hence, we have $q_j = \beta_j / (\sum_{l \neq i} \beta_l)(Q - q_i) = \beta_j / (1 - \beta_i)(Q - q_i)$ for $j \neq i$. The price can then be derived using the inverse demand function of any trader j, $I_j(q_j) = I(q_j/\beta_j) = I((Q - q_i)/(1 - \beta_i))$, where for the last transition we used our guess (6). Thus, $P_i(q_i)$, the inverse residual demand faced by trader i is given by

$$P_i(q_i) = I((Q - q_i)/(1 - \beta_i)).$$

The price impact is then given by

$$\Lambda_i = \nabla P_i(q_i) = \frac{-1}{1 - \beta_i} \nabla I((Q - q_i)/(1 - \beta_i)) = \frac{-1}{1 - \beta_i} \nabla I(Q). \tag{10}$$

Here, in the last transition, we accounted for the fact that in equilibrium, the allocation to trader i is $q_i = D_i(P) = \beta_i D(P) = \beta_i Q$.

Next, we substitute $\Lambda_i = -1/(1 - \beta_i)\nabla I(Q)$ along with $q_i = \beta_i Q$ and $I_i(q_i) = I(q_i/\beta_i) = I(Q)$ into (9). This yields the following system of partial differential equations, which governs the function $I(Q)$:

$$I(Q) - \frac{\beta_i}{1 - \beta_i} \nabla I(Q) Q = \left(\alpha_i w_0 - \beta_i Q^\top I(Q)\right) \frac{E\left[\left(Q^\top \delta\right)^{-\gamma} \delta\right]}{\beta_i E\left[\left(Q^\top \delta\right)^{1-\gamma}\right]}. \tag{11}$$

The *scale symmetry* simplifies the system of PDEs, reducing it from a set of equations involving L functions $I_i(q)$ to a system for a single function $I(Q)$.

The final simplification arises from *homogeneity*. If $I(Q)$ is a homogeneous function of degree k, then by Euler's homogeneous function theorem, we have[11]

$$\nabla I(Q) Q = k I(Q).$$

[11] The derivation is short, so we provide it here. Start with the definition of homogenous function $I(tQ) = t^k I(Q)$. Differentiate the last equation with respect to t on both sides: $\nabla I(tQ)Q = kt^{k-1} I(Q)$. Then, substitute $t = 1$.

This reduces the system of PDEs (11) to a system of linear algebraic equations:

$$\left(1 - k\frac{\beta_i}{1-\beta_i}\right) I(Q) = \left(\alpha_i w_0 - \beta_i Q^\top I(Q)\right) \frac{E\left[\left(Q^\top \delta\right)^{-\gamma}\delta\right]}{\beta_i E\left[\left(Q^\top \delta\right)^{1-\gamma}\right]}. \tag{12}$$

The system above can be solved using an approach similar to the one used in the competitive case: (1) premultiply (12) by $Q^\top$ to obtain expression for $Q^\top I(Q)$; (2) substitute the expression for $Q^\top I(Q)$ back into (12) to express $I(Q)$. We complete the derivation in the appendix, and the non-competitive equilibrium is summarized in the theorem below.

Theorem 1. *There exists a unique scale-symmetric equilibrium with a homogeneous $I(Q)$. The inverse demands are given by $I_i(q) = I(q/\beta_i)$. The function $I(q)$ is given by*

$$I(q) = \frac{w_0}{2\phi}\frac{E\left[\left(\delta^\top q\right)^{-\gamma}\delta\right]}{E\left[\left(\delta^\top q\right)^{1-\gamma}\right]}. \tag{13}$$

The scaling constants are given by

$$\beta_i = \alpha_i\phi + 1 - \sqrt{\left(\alpha_i\phi\right)^2 + 1}. \tag{14}$$

The constant ϕ is the unique positive solution to

$$\sum_{i=1}^{L}\left(\alpha_i\phi + 1 - \sqrt{\left(\alpha_i\phi\right)^2 + 1}\right) = 1. \tag{15}$$

We note several properties of our equilibrium. First, unlike in traditional linear models, our equilibrium exists even when $L = 2$. The non-existence of linear equilibrium with $L = 2$ limits the applicability of the uniform-price double auction framework in modeling financial networks, where some network regions may naturally consist of only two players.[12] Given these

[12] Malamud and Rostek (2017) and Babus and Kondor (2018) are two prominent examples of applying the uniform-price double auction to networks. The first paper effectively considers only networks with $L > 2$, while the second assumes that nodes with $L = 2$ (i.e., those with two dealers in their model) also have a mass of price-

challenges, we believe extending our approach to financial networks could provide valuable insights.

Second, we note the tractability of our model. A key challenge in solving models with market power is the price impact term $\Lambda_i(q_i)q_i$ in (9). Since price impact depends on the slopes of investors' demand functions in equilibrium, its presence in the first-order conditions transforms the problem into a system of partial differential equations (PDEs), as the FOCs relate inverse demands to the derivatives of other traders' inverse demands. In general, this system of PDEs is difficult to solve.[13]

Most of the literature circumvents this complexity by assuming that price impact is constant, as in the CARA-Normal framework, where the system of FOCs reduces to a system of algebraic equations. We propose an alternative that retains the tractability of the CARA-Normal framework while allowing for wealth effects. Specifically, we focus on settings where demands are homogeneous. By Euler's homogeneous function theorem, the price impact term $\Lambda_i(q_i)q_i$ is then proportional to the inverse demand $I_i(q_i)$ itself, which again converts the system of FOCs into a system of algebraic equations, making the model solvable while capturing wealth effects.

4.2 The cross-section of investor demands

Recall that in a scale-symmetric equilibrium, individual demands $D_i(P)$ represent a fraction β_i of the aggregate demand $D(P)$, i.e., $D_i(P) = \beta_i D(P)$. Consequently, the slope of the individual demand $D_i(P)$, which corresponds to the amount of liquidity provided by LP i, is also a fraction β_i of the slope of the aggregate demand (representing aggregate liquidity). Thus, the coefficient β_i has a dual interpretation: it represents LP i's share of total turnover and their fraction of the

taking customers. However, not all real-world networks fit these restrictions. Du and Zhu (2017) demonstrate the existence of *non-linear* equilibria in a model with linear marginal utility and $L = 2$. While this non-linear equilibrium exists, it is significantly less tractable than the linear case, which has hindered its application to network models.

[13] In some special cases, this system of PDEs can be reduced to a single ordinary differential equation (ODE) that is more tractable. See Glebkin et al. (2023a) and Glebkin et al. (2023b).

total liquidity provided. In the next proposition, we examine the cross-section of the coefficients β_i and the individual price impacts Λ_i.[14]

Proposition 2. *Larger investors have a larger share of aggregate turnover and provide more liquidity but have a higher price impact. Formally, for any i and j such that $\alpha_i > \alpha_j$, it holds that $\beta_i > \beta_j$, while $\Lambda_i > \Lambda_j$ (in the sense of positive semi-definite order).*[15] *Additionally, the share of the smallest (respectively, largest) investors in aggregate turnover exceeds (respectively, is less than) their share of aggregate wealth. Formally, ranking investors such that α_i increases with i, there exists a threshold i^* such that for any $i \geq i^*$ (respectively, $i < i^*$), $\beta_i < \alpha_i$ (respectively, $\beta_i > \alpha_i$).*

Our equilibrium exhibits several intuitive properties. First, larger investors engage in larger trades and provide more liquidity. This makes sense: with larger balance sheets, they have greater trading needs and a higher capacity to provide liquidity.

Second, consistent with Koijen and Yogo (2019), larger investors experience a greater price impact. This occurs because their price impact is determined by the liquidity provided by other investors, who, on average, contribute less liquidity.

Finally, unlike in the competitive model—where an investor's share of turnover matches their share of wealth—in the non-competitive setting, the largest investors have turnover shares smaller than their wealth shares, whereas the smallest investors exhibit the opposite pattern. This aligns with the findings of Koijen and Yogo (2019), who observe that the largest investors, managing one-third of total wealth, contribute only 4% ($< 1/3$) to the cross-sectional variance of stock returns. In contrast, the smallest investors, who also manage one-third of total wealth, account for 47% ($> 1/3$) of the cross-sectional variance of stock returns. In our model, this arises because larger investors experience a higher price impact, prompting them to adjust their demands more substantially away from the competitive benchmark—where turnover shares

[14] Note the distinction between Λ, our measure of illiquidity, which is the slope of the aggregate inverse demand, and individual price impact Λ_i, the slope of the inverse residual demand faced by LP i.

[15] That is, $\Lambda_i - \Lambda_j$ is positive definite.

align with wealth shares—compared to smaller investors.

4.3 Contrasting to competitive benchmark

In this section, we compare aggregate quantities—such as expected returns, return volatility, and liquidity—across competitive and non-competitive equilibria. Recall that quantities in the competitive equilibrium are denoted with the superscript c, while those in the non-competitive equilibrium are left without a superscript. The following proposition summarizes this comparison.

Proposition 3. *The non-competitive equilibrium is characterized by higher returns, higher return volatility, and lower illiquidity. Formally, for any k and $l \in \{1, 2, \dots, L\}$, we have*

$$\frac{\mu_k}{\mu_k^c} = \frac{\sigma_k}{\sigma_k^c} = \frac{\Lambda_{kl}^c}{\Lambda_{kl}} = \phi > 1.$$

Here, the constant ϕ is determined by (15).

As highlighted in the proposition above, when LPs exercise market power, they tilt returns (both realized and expected) in their favor, scaling them up. This scale-up results in greater volatility.[16] These results are intuitive.

A surprising result is that when traders exercise their market power, the market becomes *more* liquid. To gain intuition, consider the case of a single risky asset ($N = 1$). By comparing (9) and (8), we derive:

$$I_i(q) = I_i^c(q) - \Lambda_i(q)q.$$

This equation demonstrates that, for a given quantity $q > 0$, strategic LPs reduce their demands relative to the competitive benchmark by the amount $\Lambda_i(q)q > 0$. This behavior is

[16] The finding that greater market power is associated with higher volatility aligns with the empirical evidence presented in Ben-David et al. (2021). We discuss this connection further in the next section, where we examine the comparative statics of changes in wealth inequality, offering a more direct link to the empirical findings in that paper.

consistent with established insights on strategic behavior (see, e.g., Ausubel et al. (2014)).

In traditional linear models, where the price impact is constant, the demand reduction $\Lambda_i(q)q$ increases with q. Consequently, the slope of the inverse demand, $-I_i'(q)$, becomes larger due to the demand reduction component. This implies that LPs' demands become less elastic (since demand elasticity is inversely related to the slope of the inverse demand). As a result, traders provide less liquidity, leading to an overall decline in market liquidity.

In contrast, in our model, the demand reduction $\Lambda_i(q)q$ decreases with q (see (13), which implies that $\Lambda_i(q) \propto 1/q^2$). Consequently, the slope of the inverse demand, $-I_i'(q)$, *becomes smaller* due to the demand reduction component. This implies that demands become *more elastic*, leading to an overall *increase* in market liquidity.

We now argue that the property of demand reduction $\Lambda_i(q)q$ decreasing with q is natural. To do so, we demonstrate that if $\Lambda_i(q)q$ were to increase with q, the model would exhibit undesirable properties. Specifically, if $\Lambda_i(q)q$ increases with q, then $\lim_{q\to\infty} \Lambda_i(q)q > 0$. At the same time, $\lim_{q\to\infty} I^c(q) = 0$. Consequently, if $\Lambda_i(q)q$ increases with q, the inverse demand $I_i(q)$ would become negative for sufficiently large q. Negative prices for assets with positive payoffs would represent an arbitrage opportunity.[17]

4.4 Implications of changes in the distribution of wealth

In this section, we examine how changes in the wealth distribution, $\{\alpha_i\}_i$, influence expected returns, return volatility, and liquidity. Unlike the competitive benchmark—where variations in wealth distribution have no impact on these quantities—strategic interactions lead to a more nuanced relationship.

We focus on changes in wealth distribution that lead to an increase or decrease in inequality. When interpreting LPs as funds, an increase in inequality can result from two scenarios:

[17] This property holds more generally. In a model with symmetric CARA LPs and arbitrary payoff distributions, Glebkin et al. (2023a) demonstrate that demand reduction decreases with q for sufficiently large q. Glebkin et al. (2023b) further extend this result to accommodate general preferences.

(i) the merger of two funds or (ii) the flow of funds from a smaller fund to a larger one. In line with these scenarios, we define an increase and decrease in inequality as follows.

Definition 2. *An* ***increase in inequality*** *corresponds to the following types of changes in the wealth distribution from α to $\hat{\alpha}$:*

1. *Flow of funds from a smaller LP i to a larger LP j: $\alpha = \{\alpha_1, \ldots, \alpha_i, \ldots, \alpha_j, \ldots, \alpha_L\}$ and $\hat{\alpha} = \{\alpha_1, \ldots, \alpha_i - y, \ldots, \alpha_j + y, \ldots, \alpha_L\}$, where $y \leq \alpha_i \leq \alpha_j$.*
2. *Merger of LP i and LP j: $\alpha = \{\alpha_1, \ldots, \alpha_i, \ldots \alpha_{j-1}, \alpha_j, \alpha_{j+1}, \ldots, \alpha_L\}$ and $\hat{\alpha} = \{\alpha_1, \ldots, \alpha_i + \alpha_j, \ldots \alpha_{j-1}, \alpha_{j+1}, \ldots, \alpha_L\}$*

A ***decrease in inequality*** *corresponds to changes in the wealth distribution described in 1 and 2, but in reverse, from $\hat{\alpha}$ to α. These changes represent a flow of funds from larger to smaller LPs or the split of a single LP into two smaller entities.*

The following proposition summarizes how changes in wealth inequality influence expected returns, return volatility, and liquidity.

Proposition 4. *Consider a change in wealth distribution from α to $\hat{\alpha}$. Denote the equilibrium quantities corresponding to the distribution $\hat{\alpha}$ with a hat. We have:*

$$\frac{\hat{\mu}_k}{\mu_k} = \frac{\hat{\sigma}_k}{\sigma_k} = \frac{\Lambda_{kl}}{\hat{\Lambda}_{kl}} = \frac{\hat{\phi}}{\phi}.$$

When the change from α to $\hat{\alpha}$ corresponds to an increase (respectively, decrease) in inequality, we have $\hat{\phi} > \phi$ (respectively, $\hat{\phi} < \phi$). An increase in inequality leads to higher returns, higher return volatility, and lower illiquidity. Conversely, a decrease in inequality results in lower returns, lower return volatility, and higher illiquidity.

The results of the proposition above are intuitive: an increase in inequality amplifies the market power of LPs, producing an outcome qualitatively similar to the shift from a competitive equilibrium to a strategic one. The intuition outlined after Proposition 3 applies directly.

Our result, which shows that increased inequality leads to greater volatility, aligns with the evidence presented in Ben-David et al. (2021). They examine two scenarios involving changes in inequality: the merger of two funds (BlackRock and BGI) and an increase in the share of wealth managed by top institutions based on assets under management (AUM). In both cases, they find a positive relationship between inequality and volatility. These scenarios correspond directly to those described in Definition 2.

We provide empirical evidence supporting our surprising result that inequality and liquidity are positively related in Section 7. Notably, these results do not appear in previous models studied in the literature, as producing such an outcome requires the *combined effects of market power, wealth effects, and wealth heterogeneity.* To the best of our knowledge, our model is the first to incorporate all of these elements.

5 Is a large market competitive?

In this section, we examine the equilibrium outcomes in an economy with a large number of liquidity providers (LPs), specifically analyzing the non-competitive equilibrium in the limit as $L \to \infty$. Our focus is on determining whether aggregate quantities—such as expected returns, return volatility, and illiquidity—differ in the large non-competitive economy compared to the competitive case.

It follows from Proposition 3 that the aggregate outcomes in a large non-competitive market differ from those in a competitive market whenever $\phi(\infty) = \lim_{L\to\infty} \phi(L) > 1$. Here, $\phi(L)$ represents the constant ϕ, determined by Equation (15), in the non-competitive economy with L liquidity providers. The following proposition establishes sufficient conditions under which these outcomes differ and conditions under which they remain the same.

Proposition 5. *Expected returns, return volatility, and illiquidity in a large non-competitive market differ from those in a competitive market when the large market exhibits strictly positive* *concentration, as measured by the Herfindahl–Hirschman Index (HHI) of the wealth distribution. Conversely, if the large market has zero concentration, the aggregate outcomes in the large non-competitive market are identical to those in the competitive market. Formally, let* $\text{HHI}(L)$ *denote the Herfindahl–Hirschman Index of the wealth distribution, defined as* $\text{HHI}(L) \equiv \sum_{i=1}^{L} \alpha_i^2$*, and let* $\text{HHI}(\infty) \equiv \lim_{L\to\infty} \text{HHI}(L)$*. Assume that both* $\text{HHI}(\infty)$ *and* $\phi(\infty)$ *exist. If* $\text{HHI}(\infty) > 0$*, then* $\phi(\infty) > 1$*. Conversely, if* $\text{HHI}(\infty) = 0$*, then* $\phi(\infty) = 1$*. Moreover, we have* $\phi(\infty) \geq 1 + \text{HHI}/(1 + \sqrt{2})$.

The proposition above provides new insights into the classic question of whether markets become perfectly competitive in the limit as the number of traders approaches infinity (see, e.g., Lee and Kyle (2018) and the literature review therein). It does so by incorporating wealth effects and wealth heterogeneity, which were largely ignored in previous studies. Furthermore, the proposition establishes a connection between the wedge separating competitive and large non-competitive markets and a commonly used measure of market concentration, the Herfindahl–Hirschman Index (HHI).

According to Proposition 3, this wedge can be measured by $\phi - 1$. Indeed, the aggregate quantities in the competitive economy depend on neither L nor $\{\alpha_i\}_i$, while the quantities in the non-competitive economy can be obtained by multiplying (in the case of expected returns and volatility) or dividing (in the case of illiquidity) by ϕ.

If the HHI of a large market is non-zero, the market power of liquidity providers (LPs) persists, even as their number becomes infinite, and the wedge remains strictly positive. Conversely, if the HHI in a large market is zero, the wedge also vanishes. Thus, the HHI, a measure frequently used by the FTC to evaluate mergers, effectively captures the degree of non-competitive behavior in the market. Moreover, the final part of the proposition establishes a lower bound for this wedge as $\text{HHI}/(1 + \sqrt{2})$.[18]

We conclude with several examples.

[18] We note that in our empirical application, the lower bound described above is *tight*, as evidenced by the R^2 of regressing ϕ on HHI, which is very close to 1.

Example 1 (Homogenous LPs). *Suppose that all LPs have the same initial wealth, i.e.,*

$$\alpha_i = 1/L \text{ for all } i.$$

In this case,

$$\text{HHI}(L) = \sum_{i=1}^{L} \frac{1}{L^2} = \frac{1}{L} \to 0 \text{ as } \to \infty.$$

Thus, the wedge between competitive and large non-competitive markets becomes zero, underscoring the importance of wealth heterogeneity in generating a strictly positive wedge.

Example 2 (Power law). *Suppose the wealth distribution follows a power law:*

$$\alpha(i) = \frac{1}{\zeta(\psi)} \left(\frac{1}{i}\right)^{\psi}, \quad \psi > 1,$$

where $\zeta(\cdot)$ is the Riemann zeta function. Then,

$$\text{HHI}(\infty) = \frac{1}{\zeta(\psi)^2} \sum_{i=1}^{\infty} \left(\frac{1}{i}\right)^{2\psi} = \frac{\zeta(2\psi)}{\zeta(\psi)^2} > 0.$$

Thus, the wedge between competitive and large non-competitive markets is strictly positive.

The next example generalizes the previous one.

Example 3 (At least one granular LP). *Suppose that, in the $L \to \infty$ limit, the wealth share of at least one fund remains strictly positive. Let this fund be $i = 1$, and denote its limiting wealth share by α_1. In this case, we have:*

$$\text{HHI}(\infty) \geq \alpha_1^2 > 0.$$

Therefore, the wedge between competitive and large non-competitive markets is strictly positive.

6 Welfare

In this section, we examine the welfare implications of changes in the wealth distribution. To evaluate the welfare of LDs, we endow them with preferences, as outlined below. To maintain parsimony, the model includes only LPs and LDs without accounting for customers who might delegate investments to LPs or LDs and create fund flows. Consequently, our primary analysis focuses on changes in the wealth distribution arising from a merger or a split of funds rather than flows between funds.

6.1 Completing the model

The model studied from this point onward remains unchanged for LPs. However, we now endogenize liquidity demand by explicitly formulating the LDs' optimization problem. We assume there are $M \geq 1$ identical LDs, indexed by $m \in \{1, 2, \ldots, M\}$. Each LD is endowed with $w_0^m = w_0^{LD}/M$ units of a consumption good and a portfolio $q_0^m = q_0/M \in \mathbb{R}^{N+1}$ of assets.[19] Suppose LD m sells portfolio $q^m \in \mathbb{R}^{N+1}$ at price $P \in \mathbb{R}^{N+1}$. In equilibrium, LPs and LDs will take opposite sides of the market. Consequently, we adopt the convention that if $q_k^m > 0$, LD m *sells* q_k^m units of asset k. This contrasts with LPs, where if $q_k^i > 0$, LP i *buys* q_k^i units of asset k. LD m's time-0 consumption is given by $c_0^m = w_0^m + q^\top P$, and his time-1 consumption is $c_1^m = \delta^\top (q_0^m - q^m)$. The preferences of LDs are the same as those of LPs: they have Epstein-Zin (1989) preferences with an elasticity of intertemporal substitution (EIS) equal to 1 and a relative risk aversion (RRA) parameter equal to $\gamma \in (0, 1]$. These preferences are expressed as:

$$U_m(c_0^m, c_1^m) = \log(c_0^m) + \log E\left[(c_1^m)^{1-\gamma}\right]^{1/(1-\gamma)}.$$

[19]For simplicity, we assume that LDs are symmetric. However, our model remains tractable even if LDs are heterogeneous in terms of their initial wealth.

In contrast to LPs who submit limit orders (demand schedules), LDs submit market orders and maximize their utility while accounting for their price impact,

$$\sup_q \log\left(w_0^m + q^\top P(q)\right) + \log E\left[\left(\delta^\top (q_0 - q)\right)^{1-\gamma}\right]^{1/(1-\gamma)},$$

taking as given the equilibrium price function

$$P(q) = I(q + (M-1)q^*) = \frac{w_0}{2\phi} \frac{E\left[\left(\delta^\top (q + (M-1)q^*)\right)^{-\gamma} \delta\right]}{E\left[\left(\delta^\top (q + (M-1)q^*)\right)^{1-\gamma}\right]} \tag{16}$$

and supposing that all other LDs submit orders q^*. Price-taking LDs represent a special case, corresponding to the limit as $M \to \infty$. To ensure that equilibrium objects are well-defined, we impose the following assumption.

Assumption 2. $\delta^\top q_0 > 0$.

We assume that all model primitives are common knowledge, ensuring that LPs face no uncertainty regarding the aggregate trade of LDs. To rule out the possibility of multiple equilibria, we focus on *robust* Nash equilibria, as defined by Rostek and Weretka (2015a). In these equilibria, LPs' demand schedules remain optimal even when additive noise is introduced to their residual demand. Furthermore, we restrict our attention to equilibria with symmetric demands among LDs. Henceforth, we refer to an equilibrium satisfying these conditions as the *overall equilibrium.*

6.2 Equilibrium with endogenous liquidity demand

In this section, we characterize the overall equilibrium in our economy when liquidity demand is endogenous. This is formalized in the proposition below.

Proposition 6. *There exists a unique overall equilibrium. The inverse demand schedule of LP* *i is given by*

$$I_i(q) = I\left(\frac{q}{\beta_i}\right),$$

where $I(q)$, β_i, and ϕ are defined in (13), (14), and (15), respectively. The market order of LD m is given by $q^m = tq_0/M$, where

$$t = (1 + \eta\,(2\phi\kappa + 1))^{-1}, \quad \eta \equiv \frac{M}{M+1} \text{ and } \kappa \equiv \frac{w_0^{LD}}{w_0}. \tag{17}$$

In equilibrium, LPs' demands remain unchanged, while LDs sell a fraction $t < 1$ of their initial endowment. For the first result, since LPs' demands do not depend on the distribution of aggregate liquidity demand Q, they remain unchanged even when Q becomes endogenous. For the second result, because LPs and LDs have the same preferences, they trade the same portfolio of assets in equilibrium. Trading the portfolio q_0 helps equalize the inverse demands of LPs and LDs in equilibrium.

6.3 Benign granularity

In this section, we analyze the welfare effects of changes in the wealth distribution. Specifically, we consider the merger or split of two LPs and examine the impact of such changes on other investors. The key finding is that, under certain conditions, an increase in inequality resulting from a merger can benefit both other LPs (due to more beneficial prices) and LDs (due to improved liquidity).

Proposition 7. *Suppose that*

$$\eta < \frac{1}{2\kappa\phi\left(\sqrt{1+\bar{\alpha}^2\phi^2}\right) - 1} \quad \text{and} \quad \phi > \frac{1+\eta}{2\kappa(1-\eta)}.$$

Here, $\bar{\alpha} = \max_i \alpha_i$, and κ and η are as defined in (17). Under these conditions, a merger of two funds, i_1 and i_2, increases the utility of any LP $i \neq i_1, i_2$ and also increases the utility of *any LD. Conversely, a split of a fund into* i_1 *and* i_2 *decreases the utility of any LP* $i \neq i_1, i_2$ *and also decreases the utility of any LD.*

Now, suppose that

$$\eta > \frac{1}{2\kappa\phi\left(\sqrt{1+\underline{\alpha}^2\phi^2}\right) - 1} \quad \text{and} \quad \phi < \frac{1+\eta}{2\kappa(1-\eta)}.$$

Here, $\underline{\alpha} = \min_i \alpha_i$*, and* κ *and* η *are as defined in* (17)*. Under these conditions, a merger of two funds,* i_1 *and* i_2*, decreases the utility of any LP* $i \neq i_1, i_2$ *and also decreases the utility of any LD. Conversely, a split of a fund into* i_1 *and* i_2 *increases the utility of any LP* $i \neq i_1, i_2$ *and also increases the utility of any LD.*

The surprising result that a merger of two investors can benefit all other market participants holds when η is sufficiently small (indicating that LDs are non-competitive) and ϕ is sufficiently large (indicating that LPs' wealth is concentrated, rendering them non-competitive as well). The key mechanism underlying the beneficial effects for LDs is the impact of the merger on liquidity, as discussed in Section 4.4. Additionally, LPs benefit from improved prices, which are tilted in their favor due to increased concentration.

To understand why this result holds in a sufficiently non-competitive market, note that in such markets, strategic considerations amplify the effects of liquidity changes on market participants' behavior. The increase in liquidity resulting from the merger improves risk-sharing among investors and facilitates smoother consumption over time, ultimately leading to higher overall welfare.

Conversely, in a sufficiently competitive market, conventional wisdom holds: An increase in concentration reduces welfare by diminishing liquidity and undermining the associated benefits of risk-sharing and consumption smoothing.

7 Empirical evidence

The goal of this section is to empirically test the key predictions of our model regarding the impact of changes in wealth distribution on equilibrium prices. Specifically, Proposition 4 predicts that an increase in wealth inequality is associated with higher returns and volatility, while reducing illiquidity.

To measure wealth inequality, we use the Center for Research in Security Prices (CRSP) Mutual Funds dataset, focusing on equity mutual funds. Consistent with our theory, we capture the dynamics of wealth inequality among mutual funds by computing the changes in the HHI index ΔHHI_t of their NAV distribution.

We use VIX data from CBOE as a proxy for aggregate market volatility, and we use market capitalization-weights aggregate Amihud's lambda λ_t constructed from single stocks. We also use CRSP Value weighted index returns $R^M_{i,t}$ as the proxy for market returns. We then aggregate all data to monthly frequency and run the following regressions:

$$
\begin{aligned}
R^M_{t+1} &= \alpha + \beta^{HHI}\Delta HHI_t + \beta^{VIX} VIX_{t+1} + \beta^{lVIX} VIX_t + \beta^{Amihud}\lambda_{t+1} + \beta^{lAmihud}\lambda_t + \varepsilon_{t+1} \\
VIX_{t+1} &= \alpha + \beta^{HHI}\Delta HHI_t + \beta^{lVIX} VIX_t + \beta^{Amihud}\lambda_{t+1} + \beta^{lAmihud}\lambda_t + \varepsilon_{t+1} \\
\lambda_{t+1} &= \alpha + \beta^{HHI}\Delta HHI_t + \beta^{VIX} VIX_{t+1} + \beta^{lVIX} VIX_t + \beta^{Amihud}\lambda_t + \varepsilon_{t+1}\,.
\end{aligned}
\tag{18}
$$

We include a large number of controls in our regressions to make sure we take into account the key confounding effects; most importantly, we control for Amihud lambda and its lag and do the same for VIX because both are highly persistent. As we can see from Table 1, the results for VIX and Amihud λ are consistent with our theoretical predictions: VIX is positively related to changes in HHI, while Amihud λ is negatively related to changes in HHI. The results for the market are insignificant.[20]

[20]We have experimented with removing the various controls in our regressions, and the results for VIX and λ are largely insensitive to these choices.

Table 1: Regression Results for (18) Standard errors are heteroscedasticity-adjusted with 10 lags.

Variable	Amihud	VIX	M
const	0.0* (0.0)	2.7*** (0.7)	0.0*** (0.0)
HHI	-2.9* (1.6)	775.1** (329.7)	-2.7 (4.6)
VIX	0.0 (0.0)		-0.0*** (0.0)
lVIX	0.0 (0.0)	0.8*** (0.0)	0.0*** (0.0)
Amihud		15.9 (12.7)	-0.3*** (0.1)
lAmihud	0.5*** (0.1)	8.1 (16.0)	0.2* (0.1)
lM			-0.1** (0.1)

8 Relation to the literature

Our paper contributes to the extensive literature on strategic trading and price impact. In our model, information is symmetric, and price impact arises due to traders' limited risk-bearing capacity. We model trade using the classic double auction protocol, where traders submit price-contingent demand schedules. See, for example, Wilson (1979), Klemperer and Meyer (1989), Kyle (1989), Vayanos and Vila (1999), Vives (2011), Rostek and Weretka (2012), Kyle, Obizhaeva, and Wang (2017), Ausubel et al. (2014), Bergemann, Heumann, and Morris (2015), and Du and Zhu (2017) for the single-asset case, as well as Rostek and Weretka (2015b) and Malamud and Rostek (2017) for the multi-asset case.

These papers typically rely on the standard CARA-Normal assumption to derive linear equilibria, where the slopes of demand schedules remain independent of price levels, and equilibrium price impact (given by the inverse slope of residual supply) is constant, independent of trade size. The linearity of equilibrium critically depends on the CARA-Normal assumption, which ensures that the marginal value of asset holdings remains constant, thereby guaranteeing the existence of linear equilibria.[21]

[21] The only exception is the two-agent case, where linear equilibria fail to exist, but as shown by Du and Zhu (2017), non-linear equilibria often arise. There is also a vast literature on competitive noisy rational expectations equilibria (REE) that extends beyond the CARA-Normal framework while assuming a continuum of non-strategic traders. For instance, some papers relax the assumption of normal payoff distributions while

We offer a tractable alternative to the CARA-Normal framework. Whereas CARA-Normal models achieve tractability through the linearity of equilibria—resulting in constant price impacts—we achieve tractability by generating homogeneous equilibrium demands. This approach enables the study of wealth effects while accommodating general wealth distributions, all while preserving analytical tractability.

The double auction model in our paper enables the study of strategic liquidity provision while accounting for wealth effects. For the first time in the literature, we solve a fully micro-founded model that explicitly links market liquidity (price impact) with funding liquidity (the capital of strategic traders). Our model provides a fresh perspective on the classical results of Brunnermeier and Pedersen (2009), revealing subtle and unexpected interactions between the two forms of liquidity. In particular, we show that a higher concentration of funding liquidity can *improve* market liquidity.

There is a growing literature highlighting the significance of institutional investors in modern financial markets. Allen (2001) argues that financial crises are associated with liquidity shortages and emphasize that liquidity's effect on asset prices should be endogenous. Basak and Pavlova (2013) examine how institutional investors' trading impacts asset prices when their performance is measured relative to an index, leading to excess correlation among index stocks, heightened index stock volatility, and increased aggregate volatility. Brunnermeier and Pedersen (2009) show that institutional investors' aggregate capital (funding liquidity) influences risk premiums; see also Adrian, Etula, and Muir (2014) and He, Kelly, and Manela (2017). Micro-level evidence on individual institutional trades (Çötelioğlu, Franzoni, and Plazzi (2021), Ben-David et al. (2021)) suggests that aggregate measures overlook key market dynamics, and that investor granularity and strategic behavior—specifically, their internalization of price im-

maintaining the CARA assumption or assuming risk neutrality—see Gennotte and Leland (1990), Ausubel (1990a), Ausubel (1990b), Barlevy and Veronesi (2003), Bagnoli, Viswanathan, and Holden (2001), Yuan (2005), Breon-Drish (2015), Pálvölgyi and Venter (2015), and Chabakauri, Yuan, and Zachariadis (2017). These studies assume CARA utilities and do not incorporate wealth effects. Glebkin, Gondhi, and Kuong (2021), however, introduce wealth effects in a CARA framework by considering margin constraints whose tightness depends on wealth levels. Meanwhile, Peress (2003), Malamud (2015), and Avdis and Glebkin (2023) analyze competitive models with asymmetric information and non-CARA preferences.

pact—are crucial for understanding the interplay between market and funding liquidity. We believe our model provides a tractable framework for analyzing this link and deepening our understanding of the precise role of institutional investor granularity in asset pricing.

Our paper is part of the broad literature on the effects of illiquidity in financial markets. Many papers in this literature take market frictions as exogenous, such as constant or random trading cost, portfolio constraints, and/or assets that cannot be traded (see, Constantinides (1986), Longstaff (2009), Amihud and Mendelson (1986), Acharya and Pedersen (2005), Duffie, Gârleanu, and Pedersen (2005). In our model, the only friction is the fact that there is a finite number of large traders who behave strategically. A trader is large simply because he owns a non-negligible fraction of the aggregate wealth. Wealth effects endogenously generate (1) portfolio constraints (due to nonnegativity of wealth), (2) illiquidity due to endogenous price impact, and (3) systemic liquidity that is priced in the cross-section of asset returns.[22]

The price impact of institutional trades has been extensively documented in the literature. See, for example, Chan and Lakonishok (1995), Griffin, Harris, and Topaloglu (2003), Chiyachantana, Jain, Jiang, and Wood (2004), Almgren, Thum, Hauptmann, and Li (2005), Coval and Stafford (2007), and Ben-David et al. (2021). Notably, Chung and Huh (2016) find that price impact is a priced factor and has a stronger effect on returns than adverse selection. Focusing on the non-informational component of price impact, our model provides a general framework for analyzing the relationship between the distribution of funding liquidity and market liquidity.

While our approach relies on supply function equilibria, as in Wilson (1979), Klemperer and Meyer (1989), and Kyle (1989), a related strand of literature models imperfect competition among traders in a *Cournot* fashion, where large traders are restricted to submitting market orders. See Gabszewicz and Vial (1972), Vives (1988), and, more recently, Neuhann, Sefidgaran, and Sockin (2021) and Neuhann and Sockin (2024). Through the lens of our model,

[22] Acharya and Bisin (2014) endogenize default risk using counterparty risk when positions are opaque. While there is no counterparty risk in our model, an agent's ability to borrow from other agents is effectively limited by the amount of liquid wealth that he can post as collateral.

this approach primarily captures imperfect competition among LDs, providing complementary insights to our focus on competition among LPs. Additionally, none of these studies examine the relationship between the distribution of wealth and market liquidity, which is central to our paper.

Finally, our paper provides new insights into the classic question of whether markets become perfectly competitive as the number of traders approaches infinity (see, e.g., Wilson (1977), Satterthwaite and Williams (1989), Rustichini, Satterthwaite, and Williams (1994), and, more recently, Lee and Kyle (2018)). Unlike previous studies, we incorporate wealth effects and wealth heterogeneity, which were largely ignored in previous studies. Furthermore, we establish a link between the wedge separating competitive and large non-competitive markets and a widely used measure of market concentration, the Herfindahl-Hirschman Index (HHI).

9 Conclusion

This paper examines the impact of wealth concentration among a few large investors on asset prices and liquidity. Our findings align with empirical evidence, showing that greater concentration leads to higher volatility and returns while large investors trade less relative to their wealth share. Surprisingly, we also find that higher concentration enhances liquidity, a result supported by our empirical analysis.

Furthermore, we demonstrate that increased concentration can *improve* welfare when markets are sufficiently non-competitive. We establish a link between the wedge separating competitive and non-competitive outcomes and the Herfindahl-Hirschman Index (HHI), a standard measure of wealth concentration. Notably, this wedge can persist even in large markets, implying that market power does not necessarily vanish as the number of traders grows. Our results suggest that institutional concentration has complex effects on market functioning, challenging the conventional view that it is always detrimental.

The model can be extended in several directions. First, incorporating more general wealth distributions among LDs could shed light on the interaction between wealth concentration in both trader groups. Second, extending the model to a network setting could provide insights into how the distribution of wealth over the financial network affects asset prices and market quality. These extensions are left for future research.

Appendices

A A Summary of Notation

Notation	Explanation
	General mathematical notation
$q^\top$	Transpose of a vector q
$\nabla f(q),$ where $f: \mathbb{R}^N \to \mathbb{R}$	Gradient of f, $(\nabla f)_l = \frac{\partial f}{\partial q_l}$
$\nabla^2 f(q),$ where $f: \mathbb{R}^N \to \mathbb{R}$	Hessian of f, $(\nabla^2 f)_{kl} = \frac{\partial^2 f}{\partial q_k \partial q_l}$
$\nabla I(q),$ where $I: \mathbb{R}^N \to \mathbb{R}^N$	Jacobian of I, $(\nabla f)_{ik} = \frac{\partial I^i}{\partial q_k}$
A_{ij}	ij-th element of a matrix A.
a_i	i-th element of a vector a.
	Model variables
$I^i(q)$	Trader i's inverse demand. $I^i_k(q)$ is a price that a trader i bids for asset k, given that he gets allocation q.
$P_i(q_i)$	Inverse residual demand faced by a trader i.
$\Lambda_i(q_i) = \nabla P_i(q_i)$	Price impact matrix of a trader i.

Notation	Explanation
β_i	Scaling constants. We have $I_i(q) = I(q/\beta_i)$ in a scale-symmetric equilibrium
α_i	Investor i's share of the total wealth.
$\mathrm{HHI} = \sum \alpha_i^2$	Herfindahl-Hirschman Index (HHI) of the wealth distribution.
$\eta = \frac{M}{M+1}$	Index characterizing the competitiveness of LDs, where M is the number of LDs.
$\kappa = \frac{w_0^{LD}}{w_0}$	Ratio of aggregate wealths of LDs to LPs.

B Proofs

B.1 Proof of Lemma 1

Lemma 1. *The function* $f(q) = \frac{E\left[\left(q^\top \delta\right)^{-\gamma}\delta\right]}{E\left[\left(q^\top \delta\right)^{1-\gamma}\right]}$ *is strictly decreasing in* q.

Proof. Note that $f(q)$ can be written as $f(q) = \nabla g(q)$, where $g(q) = \log\left(E\left[\left(q^\top \delta\right)^{1-\gamma}\right]^{\frac{1}{1-\gamma}}\right)$. Since $g(q)$ is concave, its' gradient is strictly decreasing. ■

B.2 Proof of Theorem 1

Proof of Theorem 1. As derived in Section 4.1, the inverse residual demand that trader i faces is given by

$$I\left(\frac{Q - q^i}{1 - \beta_i}\right),$$

where q^i represents the portfolio trader i intends to trade, and Q denotes a specific realization of supply. Therefore, trader i's ex-post optimization problem can be written as

$$\sup_q \left\{ \log \left(\alpha_i w_0 - q^\top I \left(\frac{Q-q}{1-\beta_i} \right) \right) + \log \left(E \left[\left(q^\top \delta \right)^{1-\gamma} \right]^{\frac{1}{1-\gamma}} \right) \right\}. \tag{19}$$

The first-order condition yields

$$I \left(\frac{Q-q}{1-\beta_i} \right) - \frac{1}{1-\beta_i} \nabla I \left(\frac{Q-q}{1-\beta_i} \right) q = \left(\alpha_i w_0 - q^\top I \left(\frac{Q-q}{1-\beta_i} \right) \right) \frac{E \left[\left(q^\top \delta \right)^{-\gamma} \delta \right]}{E \left[\left(q^\top \delta \right)^{1-\gamma} \right]}. \tag{20}$$

Lemma 2 below establishes that the optimization problem (19) is globally concave. Consequently, the first-order condition (20) is both necessary and sufficient.

In the scale-symmetric equilibrium, $q = \beta_i Q$ must be optimal for any admissible Q. Substituting $q = \beta_i Q$ into the expression above yields the system of PDEs (11). Applying homogeneity then reduces this system to the algebraic equations (12). For convenience, we restate (12) below:

$$\left(1 - k \frac{\beta_i}{1-\beta_i} \right) I(Q) = \left(\alpha_i w_0 - \beta_i Q^\top I(Q) \right) \frac{E \left[\left(Q^\top \delta \right)^{-\gamma} \delta \right]}{\beta_i E \left[\left(Q^\top \delta \right)^{1-\gamma} \right]}. \tag{21}$$

Multiply by $Q^\top$ to find $E = Q^\top I(Q)$ is a constant that solves

$$E = \left(\alpha_i w_0 - \beta_i E \right) 1/\beta_i + \frac{k \beta_i}{1-\beta_i} E. \tag{22}$$

Then it follows from (21) that $I_i(q) \propto \frac{E\left[\left(\delta^\top q \right)^{-\gamma} \delta \right]}{E\left[\left(\delta^\top q \right)^{1-\gamma} \right]}$ and so the demand function is homogenous of degree -1. Thus, $k = -1$. Substituting $k = -1$ back to (22), we obtain

$$E = \frac{\alpha_i (1-\beta_i) w_0}{(2-\beta_i)\beta_i}.$$

And, from (21) we obtain

$$I(q) = \frac{\alpha_i(1-\beta_i)w_0}{(2-\beta_i)\beta_i} \frac{E\left[\left(\delta^\top q\right)^{-\gamma}\delta\right]}{E\left[\left(\delta^\top q\right)^{1-\gamma}\right]}.$$

For the scale-symmetric equilibrium to exist, we must have that

$$\frac{\alpha_i(1-\beta_i)}{(2-\beta_i)\beta_i} = \frac{1}{2\phi}, \tag{23}$$

for some constant ϕ. There is a unique solution to (23) which is between 0 and 1, given by

$$\beta_i = \alpha_i\phi + 1 - \sqrt{(\alpha_i\phi)^2 + 1}.$$

The constant ϕ is pinned down by the condition $\sum_i \beta_i = 1$:

$$\sum_i \left(\alpha_i\phi + 1 - \sqrt{(\alpha_i\phi)^2 + 1}\right) = 1$$

The solution to the equation above exists, as the function on the left-hand side is continuous and attains 0 at $\phi \to 0$ and goes to $L > 1$ as $\phi \to \infty$. The solution is unique as the function is monotone. ■

Lemma 2. *The optimization problem (19) is concave on the set of admissible portfolios when $\gamma \leq 1$.*

Proof. It is sufficient to show that the function $f(q) = q^\top I(Q-q)$ is convex, where $I(q)$ is defined by (13). To establish this, it suffices to demonstrate that $g(y) = \frac{E\left[\left(\delta^\top y\right)^{-\gamma}\delta^\top Q\right]}{E\left[\left(\delta^\top y\right)^{1-\gamma}\right]}$ is convex in y. (Indeed, rearranging (13), one can obtain $g(Q-q) - 2\phi/w_0\, f(q) = 1$.) We compute the

Hessian of $g(y)$:

$$\nabla^2 g(y) =$$

$$\gamma(\gamma+1)\frac{E\left[\left(\delta^\top y\right)^{-\gamma-2}\left(\delta^\top Q\right)\delta\delta^\top\right]}{E\left[\left(\delta^\top y\right)^{1-\gamma}\right]} + 2\gamma(1-\gamma)\frac{E\left[\left(\delta^\top y\right)^{-\gamma-1}\left(\delta^\top Q\right)\delta\right]E\left[\left(\delta^\top y\right)^{-\gamma}\delta^\top\right]}{E\left[\left(\delta^\top y\right)^{1-\gamma}\right]^2}$$
$$+\gamma(1-\gamma)\frac{E\left[\left(\delta^\top y\right)^{-\gamma}\delta^\top Q\right]E\left[\left(\delta^\top y\right)^{-\gamma-1}\delta\delta^\top\right]}{E\left[\left(\delta^\top y\right)^{1-\gamma}\right]^2}$$
$$+2(1-\gamma)^2\frac{E\left[\left(\delta^\top y\right)^{-\gamma}\delta^\top Q\right]E\left[\left(\delta^\top y\right)^{-\gamma}\delta\right]E\left[\left(\delta^\top y\right)^{-\gamma}\delta^\top\right]}{E\left[\left(\delta^\top y\right)^{1-\gamma}\right]^{1+\sqrt{2}}}.$$

For any admissible portfolio x, we have that $x^\top \nabla^2 g\, x > 0$. To see why, consider, for example, the second term in the last displayed equation. After premultiplying by $x^\top$ and multiplying by x, it will yield

$$\frac{E\left[\left(\delta^\top y\right)^{-\gamma-1}\left(\delta^\top Q\right)\delta^\top x\right]E\left[\left(\delta^\top y\right)^{-\gamma}\delta^\top x\right]}{E\left[\left(\delta^\top y\right)^{1-\gamma}\right]^2},$$

which is positive as $\delta^\top x > 0$ for an admissible x. ■

B.3 Proof of Proposition 2

Proof of Proposition 2.

The function

$$g(x;\, a) \;=\; ax + 1 - \sqrt{(ax)^2+1}$$

is strictly increasing in x, for any arbitrary constant $a > 1$. Thus, for any i and j such that $\alpha_i > \alpha_j$, we have

$$\beta_i \;=\; g(\alpha_i;\, \phi) \;>\; g(\alpha_j;\, \phi) \;=\; \beta_j.$$

.

From the relationship $\Lambda_i \propto 1/(1-\beta_i)$ (cf. (10)), it immediately follows that for any i and j such that $\alpha_i > \alpha_j$, we also have $\Lambda_i > \Lambda_j$ in the elementwise sense.

For the final part, consider $\beta_i = \alpha_i\phi + 1 - \sqrt{(\alpha_i\phi)^2+1}$ as a function of $\alpha_i \in [0,1]$ for a given ϕ. It can be shown that β_i is concave, starts at zero, and crosses the 45-degree line ($\beta_i = \alpha_i$) exactly once for $\alpha_i > 0$. Consequently, there can be at most one threshold i^*.

Such a threshold must exist because the largest β_i is smaller than the largest α_i. If this were not the case, then given the single crossing property established earlier, we would have $\beta_i > \alpha_i$ for all i, violating the condition to pin down ϕ in (15). Furthermore, there must exist some i for which $\beta_i > \alpha_i$, as otherwise (15) would again be violated. ■

B.4 Proof of Proposition 3

Proof of Proposition 3. It suffices to prove that, in equilibrium, $\phi > 1$. Consider (15). The left-hand side of this equation is a continuously increasing function of ϕ that approaches $L > 1$ as $\phi \to \infty$. Therefore, it is enough to show that the left-hand side of (15) is strictly less than 1 when $\phi = 1$.

Indeed, by multiplying and dividing by $\alpha_i\phi + 1 + \sqrt{(\alpha_i\phi)^2+1}$, we can rewrite the left-hand side of (15) as

$$\sum_i \frac{2\alpha_i^2\phi^2}{\alpha_i\phi + 1 + \sqrt{(\alpha_i\phi)^2+1}}\Bigg|_{\phi=1} = \sum_i \frac{2\alpha_i^2}{\alpha_i + 1 + \sqrt{\alpha_i^2+1}}.$$

This sum satisfies

$$\sum_i \frac{2\alpha_i^2}{\alpha_i + 1 + \sqrt{\alpha_i^2+1}} < \sum_i \alpha_i^2 < \sum_i \alpha_i = 1.$$

■

B.5 Proof of Proposition 4

Proof of Proposition 4. The relationship $\frac{\hat{\mu}_k}{\mu_k} = \frac{\hat{\sigma}_k}{\sigma_k} = \frac{\hat{\Lambda}_{kl}}{\Lambda_{kl}} = \frac{\hat{\phi}}{\phi}$ follows directly from Proposition 3. For instance, for expected returns, we can derive:

$$\frac{\hat{\mu}_k}{\mu_k} = \frac{\hat{\mu}_k}{\hat{\mu}_k^c} \cdot \frac{\mu_k^c}{\mu_k} = \frac{\hat{\phi}}{\phi}.$$

To show that an increase in inequality leads to an increase in ϕ, we analyze the impact of wealth redistribution. The case of a decrease in inequality is analogous and omitted for brevity.

Consider a change in the wealth distribution from α to $\hat{\alpha}$, corresponding to a transfer of funds from a smaller LP i to a larger LP j. Specifically,

$$\alpha = \{\alpha_1, \ldots, \alpha_i, \ldots, \alpha_j, \ldots, \alpha_L\}, \quad \hat{\alpha} = \{\alpha_1, \ldots, \alpha_i - y, \ldots, \alpha_j + y, \ldots, \alpha_L\},$$

where $y \leq \alpha_i \leq \alpha_j$.

Define $b(\alpha, \phi) = \alpha\phi + 1 - \sqrt{(\alpha\phi)^2 + 1}$. The equation (15), which determines ϕ, can then be expressed as:

$$\sum_i b(\alpha_i, \phi) = 1.$$

Observe that for a given ϕ, the function $b(\alpha, \phi)$ is concave and increasing in α. Consequently,

$$|b(\alpha_i - y, \phi) - b(\alpha_i, \phi)| > b(\alpha_j + y, \phi) - b(\alpha_j, \phi).$$

Thus, for a given ϕ,

$$\sum_i b(\alpha_i, \phi) > \sum_i b(\hat{\alpha}_i, \phi).$$

Since $b(\alpha, \phi)$ is increasing in ϕ, for the equation $\sum_i b(\hat{\alpha}_i, \hat{\phi}) = 1$ to hold, we must have $\hat{\phi} > \phi$.

The case of a merger is equivalent (in terms of pinning down ϕ) to the case considered above with $y = \alpha_i$. ■

B.6 Proof of Proposition 5

Proof of Proposition 5. Note that we can rewrite the left-hand side of (15) as

$$\phi + \sum_i \left(1 - \sqrt{(\alpha_i \phi)^2 + 1}\right) = \phi - \sum_i \frac{\alpha_i^2 \phi^2}{1 + \sqrt{(\alpha_i \phi)^2 + 1}}. \tag{24}$$

The last step follows by multiplying and dividing each term in the sum by $1 + \sqrt{(\alpha_i \phi)^2 + 1}$, and then applying the identity $(a - b)(a + b) = a^2 - b^2$.

Now, note that since $\alpha_i < 1$ for all i, we can write

$$\phi - \sum_i \frac{\alpha_i^2 \phi^2}{1 + \sqrt{(\alpha_i \phi)^2 + 1}} \leq \phi - \sum_i \frac{\alpha_i^2 \phi^2}{1 + \sqrt{\phi^2 + 1}} = \phi - \frac{\text{HHI}\, \phi^2}{1 + \sqrt{\phi^2 + 1}}.$$

Moreover, since $\frac{\text{HHI}\, \phi^2}{1+\sqrt{\phi^2+1}}$ is a strictly increasing function of ϕ, and in equilibrium $\phi \geq 1$ (see the proof of Proposition 3), we can further write

$$\phi - \frac{\text{HHI}\, \phi^2}{1 + \sqrt{\phi^2 + 1}} \leq \phi - \frac{\text{HHI}}{1 + \sqrt{2}}. \tag{25}$$

Recall that $\phi(L)$ solves the equation

$$\phi - \sum_i \frac{\alpha_i^2 \phi^2}{1 + \sqrt{(\alpha_i \phi)^2 + 1}} = 1.$$

From the inequality (25), it follows that $\phi(L)$ is greater than the solution to $\phi - \frac{\text{HHI}}{1+\sqrt{2}} = 1$. Hence,

$$\phi(L) \geq 1 + \frac{\text{HHI}(L)}{1 + \sqrt{2}}.$$

Taking the limit as $L \to \infty$, we obtain

$$\phi(\infty) \geq 1 + \frac{\text{HHI}(\infty)}{1+\sqrt{2}}.$$

Thus, if $\text{HHI}(\infty) > 0$, it follows that $\phi(\infty) > 1$.

We now turn to deriving an upper bound for ϕ. Consider the right-hand side of (24), and note that since $\alpha_i \geq 0$ for all i, we can write

$$\phi - \sum_i \frac{\alpha_i^2 \phi^2}{1+\sqrt{(\alpha_i \phi)^2 + 1}} \geq \phi - \sum_i \frac{\alpha_i^2 \phi^2}{2} = \phi - \phi^2 \frac{\text{HHI}}{2}.$$

From this inequality, it follows that $\phi(L)$ is less than the smaller solution to $\phi - \phi^2 \frac{\text{HHI}}{2} = 1$. Denote this solution by $\phi^*(L)$. Hence,

$$1 \leq \phi(\infty) \leq \phi^*(\infty).$$

The left-hand side of the above inequality follows because in equilibrium $\phi(L) \geq 1$.

Since $\phi^*(\infty) = 1$ when $\text{HHI}(\infty) = 0$, we conclude that if $\text{HHI}(\infty) = 0$, then $\phi(\infty) = 1$.

■

B.7 Proof of Proposition 6

Proof of Proposition 6. The derivation and characterization of LPs' demand schedules remain unchanged from Theorem 1.

Now, consider LD m and suppose that all other LDs submit orders q^*. If the LD of interest trades a quantity q, the resulting price will be given by

$$P(q) = I(q + (M-1)q^*) = \frac{w_0}{2\phi} \frac{E\left[\left(\delta^\top (q + (M-1)q^*)\right)^{-\gamma} \delta\right]}{E\left[\left(\delta^\top (q + (M-1)q^*)\right)^{1-\gamma}\right]}. \tag{26}$$

The price impact is given by

$$\Lambda^m(q) = -\nabla P(q) = \frac{w_0}{2\phi}\Bigg(\gamma \frac{E\left[\left(\delta^\top (q + (M-1)q^*)\right)^{-\gamma-1} \delta\delta^\top\right]}{E\left[\left(\delta^\top (q + (M-1)q^*)\right)^{1-\gamma}\right]} + (1-\gamma)\frac{E\left[\left(\delta^\top (q + (M-1)q^*)\right)^{-\gamma} \delta\right] E\left[\left(\delta^\top (q + (M-1)q^*)\right)^{-\gamma} \delta^\top\right]}{E\left[\left(\delta^\top (q + (M-1)q^*)\right)^{1-\gamma}\right]^2}\Bigg). \tag{27}$$

LD m solves the following problem

$$\sup_q \log\left(w_0^m + q^\top P(q)\right) + \log E\left[\left(\delta^\top (q_0 - q)\right)^{1-\gamma}\right]^{1/(1-\gamma)}.$$

The first-order necessary and sufficient condition is given by:[23]

$$P(q) = \left(w_0^m + q^\top P(q)\right) \frac{E\left[\left(\delta^\top (q_0^m - q)\right)^{-\gamma} \delta\right]}{E\left[\left(\delta^\top (q_0^m - q)\right)^{1-\gamma}\right]} - \Lambda^m(q)q.$$

Given the symmetry, $q = q^*$ must satisfy the FOC above. Substituting this into the equation and denoting $Q^* = Mq^*$, while accounting for the relationships $\Lambda^m(q^*)Q^* = I(Q^*)$ and $P(q^*) = I(Q^*)$ (they follow from (26) and (27)) as well as $q_0^m = q_0/M$ and $w_0^m = w_0^{LD}/M$, we obtain:

$$I(Q^*) = \left(w_0^{LD} + Q^* I(Q^*)\right) \frac{E\left[\left(\delta^\top (q_0 - Q^*)\right)^{-\gamma} \delta\right]}{E\left[\left(\delta^\top (q_0 - Q^*)\right)^{1-\gamma}\right]} - \frac{I(Q^*)}{M}.$$

[23]The proof of sufficiency follows the same reasoning as in Theorem 1, using Lemma 2.

Moreover, it holds that $Q^* I(Q^*) = w_0/(2\phi)$. Therefore,

$$\frac{M+1}{M}\frac{w_0}{2\phi}\frac{E\left[\left(\delta^\top Q^*\right)^{-\gamma}\delta\right]}{E\left[\left(\delta^\top Q^*\right)^{1-\gamma}\right]} = \left(w_0^{LD} + \frac{w_0}{2\phi}\right)\frac{E\left[\left(\delta^\top\left(q_0 - Q^*\right)\right)^{-\gamma}\delta\right]}{E\left[\left(\delta^\top\left(q_0 - Q^*\right)\right)^{1-\gamma}\right]}. \tag{28}$$

We guess that the solution to the above equation is of the form $Q^* = tq_0$ for some scalar t. (Lemma 3 below shows that no other solutions exist.) Then, t solves

$$\frac{M+1}{M}\frac{w_0}{2\phi}\frac{1}{t} = \left(w_0^{LD} + \frac{w_0}{2\phi}\right)\frac{1}{1-t} \iff t = \left(1 + \frac{M}{M+1}\left(2\phi\frac{w_0^{LD}}{w_0} + 1\right)\right)^{-1}.$$

Note that since $t < 1$, the equilibrium objects are well-defined.

■

Lemma 3. *There is a unique solution to* (28).

Proof of Lemma 3. Denote

$$k = \left(w_0^{LD} + \frac{w_0}{2\phi}\right)\frac{M}{M+1}\frac{2\phi}{w_0}, \quad \text{and} \quad f(q) = \frac{E\left[\left(\delta^\top q\right)^{-\gamma}\delta\right]}{E\left[\left(\delta^\top q\right)^{1-\gamma}\right]}.$$

Then, equation (28) can be rewritten as:

$$f(Q^*) = kf(q_0 - Q^*).$$

Assume, for the sake of contradiction, that there exist two distinct solutions $q_1 \neq q_2$ such that:

$$f(q_1) = kf(q_0 - q_1) \quad \text{and} \quad f(q_2) = kf(q_0 - q_2). \tag{29}$$

Since $f(q)$ is strictly decreasing in q (by Lemma 1), we have:

$$(q_2 - q_1)^\top (f(q_2) - f(q_1)) < 0. \tag{30}$$

Similarly, we have

$$(q_0 - q_2 - (q_0 - q_1))^\top (f(q_0 - q_2) - f(q_0 - q_1)) = (q_1 - q_2)^\top (f(q_0 - q_2) - f(q_0 - q_1)) < 0. \quad (31)$$

However, substituting $f(q_1) = kf(q_0 - q_1)$ and $f(q_2) = kf(q_0 - q_2)$ from (29) into (30), we get:

$$k(q_2 - q_1)^\top (f(q_0 - q_2) - f(q_0 - q_1)) < 0. \quad (32)$$

The inequalities (31) and (32) create a contradiction. ■

B.8 Proof of Proposition 7

Proof of Proposition 7.

Consider LP i. Their equilibrium utility is given by:

$$U_i = \log(\alpha_i w_0 - (q_i^*)^\top P_*) + \log\left(E\left[\left((q_i^*)^\top \delta\right)^{1-\gamma}\right]^{\frac{1}{1-\gamma}}\right). \quad (33)$$

The equilibrium allocations and prices are:

$$q_i = \beta_i Q_*, \quad P_* = I(Q_*), \quad Q_* = tq_0,$$

and we have $Q_*^\top I(Q_*) = w_0/(2\phi)$. Substituting these expressions into (33), we obtain:

$$\begin{aligned} U_i &= \log\left(\alpha_i w_0 - \beta_i \frac{w_0}{2\phi}\right) + \log\left(\beta_i t E\left[\left(q_0^\top \delta\right)^{1-\gamma}\right]^{\frac{1}{1-\gamma}}\right) \\ &= \log\left(\alpha_i - \frac{\beta_i}{2\phi}\right) + \log(\beta_i t) + \cdots \end{aligned} \quad (34)$$

Here, the terms denoted by $\cdots$ are unaffected by changes in the wealth distribution.

Now consider LD m. Their equilibrium utility is:

$$U_m = \log\left(w_0^m + (q_*^m)^\top P_*\right) + \log\left(E\left[\left((q_0^m - q_*^m)^\top \delta\right)^{1-\gamma}\right]^{\frac{1}{1-\gamma}}\right). \tag{35}$$

The equilibrium allocations and prices are:

$$q_*^m = \frac{Q_*}{M}, \quad P_* = I(Q_*), \quad Q_* = tq_0, \quad w_0^m = \frac{w_0^{LD}}{M}, \quad q_0^m = \frac{q_0}{M},$$

and we have $Q_*^\top I(Q_*) = w_0/(2\phi)$. Substituting these expressions into (35), we obtain:

$$U_m = \log\left(\frac{w_0^{LD}}{M} + \frac{w_0}{2\phi}\right) + \log(1-t) + \cdots$$

Here again, the terms denoted by $\cdots$ are unaffected by changes in the wealth distribution.

Substituting the expressions for t from (17) and β_i from (14) into the utility expressions U_i and U_m ((34) and (35), respectively) and differentiating the resulting expressions with respect to ϕ, we find that U_i is increasing in ϕ if and only if:

$$\frac{1}{\phi\sqrt{1+\alpha_i^2\phi^2}} - \frac{2\eta\kappa}{1+\eta+2\eta\kappa\phi} > 0. \tag{36}$$

Here, the notations κ and η are defined in (17).

Similarly, U_m is increasing in ϕ if and only if:

$$-\frac{1}{\phi} + \frac{4\kappa}{1+2\kappa\phi} - \frac{2\eta\kappa}{1+\eta+2\eta\kappa\phi} > 0. \tag{37}$$

Since ϕ is the only input to U_i and U_m that is affected by a merger or split (for LP i not participating in the merger or the split), and ϕ increases after the merger and decreases after the split, the results directly imply the statement of the Proposition.

Indeed, for the first statement, note that since $\alpha_i < \max_i \alpha_i = \bar{\alpha}$, inequality (36) holds if

$$\frac{1}{\phi\sqrt{1+\bar{\alpha}^2\phi^2}} > \frac{2\eta\kappa}{1+\eta+2\eta\kappa\phi},$$

which is equivalent to:

$$\kappa < \frac{1+\eta}{2\eta\phi(\sqrt{1+\bar{\alpha}^2\phi^2}-1)}.$$

Similarly, inequality (37) holds if:

$$\phi > \frac{1+\eta}{2\kappa(1-\eta)}.$$

The second statement is proved analogously by noting that $\alpha_i > \min_i \alpha_i = \underline{\alpha}$.

■

References

Viral Acharya and Alberto Bisin. Counterparty risk externality: Centralized versus over-the-counter markets. Journal of Economic Theory, 149:153–182, 2014.

Viral V. Acharya and Lasse Heje Pedersen. Asset pricing with liquidity risk. Journal of Financial Economics, 77(2):375–410, 2005.

Tobias Adrian, Erkko Etula, and Tyler Muir. Financial intermediaries and the cross-section of asset returns. The Journal of Finance, 69(6):2557–2596, 2014.

Franklin Allen. Do financial institutions matter? The Journal of Finance, 56(4):1165–1175, 2001.

Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Direct estimation of equity market impact. Risk, 18(7):58–62, 2005.

Yakov Amihud. Illiquidity and stock returns: cross-section and time-series effects. Journal of financial markets, 5(1):31–56, 2002.

Yakov Amihud and Haim Mendelson. Asset pricing and the bid-ask spread. Journal of financial Economics, 17(2):223–249, 1986.

Lawrence M Ausubel. Insider trading in a rational expectations economy. The American Economic Review, pages 1022–1041, 1990a.

Lawrence M Ausubel. Partially-revealing rational expectations equilibrium in a competitive economy. Journal of Economic Theory, 50(1):93–126, 1990b.

Lawrence M Ausubel, Peter Cramton, Marek Pycia, Marzena Rostek, and Marek Weretka. Demand reduction and inefficiency in multi-unit auctions. The Review of Economic Studies, 81(4):1366–1400, 2014.

Efstathios Avdis and Sergei Glebkin. Chile. Working Paper, 2023.

Ana Babus and Péter Kondor. Trading and information diffusion in over-the-counter markets. Econometrica, 86(5):1727–1769, 2018.

Mark Bagnoli, S Viswanathan, and Craig Holden. On the existence of linear equilibria in models of market making. Mathematical Finance, 11(1):1–31, 2001.

Gadi Barlevy and Pietro Veronesi. Rational panics and stock market crashes. Journal of Economic Theory, 110(2):234–263, 2003.

Suleyman Basak and Anna Pavlova. Asset prices and institutional investors. American Economic Review, 103(5):1728–1758, 2013.

Itzhak Ben-David, Francesco Franzoni, Rabih Moussawi, and John Sedunov. The granular nature of large institutional investors. Management Science, 67(11):6629–6659, 2021.

Dirk Bergemann, Tibor Heumann, and Stephen Morris. Information and volatility. Journal of Economic Theory, 158:427–465, 2015.

Bradyn Breon-Drish. On existence and uniqueness of equilibrium in a class of noisy rational expectations models. The Review of Economic Studies, 82(3):868–921, 2015.

Markus K Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. The review of financial studies, 22(6):2201–2238, 2009.

Georgy Chabakauri, Kathy Yuan, and Konstantinos E Zachariadis. Multi-asset noisy rational expectations equilibrium with contingent claims. LSE working paper, 2017.

Louis KC Chan and Josef Lakonishok. The behavior of stock prices around institutional trades. The Journal of Finance, 50(4):1147–1174, 1995.

Chiraphol N Chiyachantana, Pankaj K Jain, Christine Jiang, and Robert A Wood. International evidence on institutional trading behavior and price impact. The Journal of Finance, 59(2): 869–898, 2004.

Kee H Chung and Sahn-Wook Huh. The noninformation cost of trading and its relative importance in asset pricing. The Review of Asset Pricing Studies, 6(2):261–302, 2016.

George M Constantinides. Capital market equilibrium with transaction costs. Journal of Political Economy, 94(4):842–862, 1986.

Efe Çötelioğlu, Francesco Franzoni, and Alberto Plazzi. What constrains liquidity provision? evidence from institutional trades. Review of Finance, 25(2):485–517, 2021.

Joshua Coval and Erik Stafford. Asset fire sales (and purchases) in equity markets. Journal of Financial Economics, 86(2):479–512, 2007.

Songzi Du and Haoxiang Zhu. Bilateral trading in divisible double auctions. Journal of Economic Theory, 167:285–311, 2017.

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

Larry G Epstein and Stanley E Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. Econometrica (1986-1998), 57(4): 937, 1989.

Jean Jaskold Gabszewicz and Jean-Philippe Vial. Oligopoly "a la cournot" in a general equilibrium analysis. Journal of economic theory, 4(3):381–400, 1972.

Gerard Gennotte and Hayne Leland. Market liquidity, hedging, and crashes. The American Economic Review, pages 999–1021, 1990.

Sergei Glebkin and John Chi-Fong Kuong. When large traders create noise. Journal of Financial Economics, 150(2):103709, 2023.

Sergei Glebkin, Naveen Gondhi, and John Chi-Fong Kuong. Funding constraints and informational efficiency. The Review of Financial Studies, 34(9):4269–4322, 2021.

Sergei Glebkin, Semyon Malamud, and Alberto Teguia. Illiquidity and higher cumulants. The Review of Financial Studies, 36(5):2131–2173, 2023a.

Sergei Glebkin, Semyon Malamud, and Alberto Teguia. Strategic trading with wealth effects. 2023b.

John M Griffin, Jeffrey H Harris, and Selim Topaloglu. The dynamics of institutional and individual trading. The Journal of Finance, 58(6):2285–2320, 2003.

Zhiguo He, Bryan Kelly, and Asaf Manela. Intermediary asset pricing: New evidence from many asset classes. Journal of Financial Economics, 126(1):1–35, 2017.

Paul D Klemperer and Margaret A Meyer. Supply function equilibria in oligopoly under uncertainty. Econometrica: Journal of the Econometric Society, pages 1243–1277, 1989.

Ralph S. J. Koijen and Motohiro Yogo. A demand system approach to asset pricing. Journal of Political Economy, 127(4):1475–1515, 2019. doi: 10.1086/701683. URL https://doi.org/10.1086/701683.

Albert S Kyle. Informed speculation with imperfect competition. The Review of Economic Studies, 56(3):317–355, 1989.

Albert S Kyle, Anna A Obizhaeva, and Yajun Wang. Smooth trading with overconfidence and market power. The Review of Economic Studies, 85(1):611–662, 2017.

Jeongmin Lee and Albert S Kyle. When are financial markets perfectly competitive? 2018.

Francis A Longstaff. Portfolio claustrophobia: Asset pricing in markets with illiquid assets. American Economic Review, 99(4):1119–1144, 2009.

Semyon Malamud. Noisy arrow-debreu equilibria. Available at SSRN 2572881, 2015.

Semyon Malamud and Marzena Rostek. Decentralized exchange. American Economic Review, 107(11):3320–62, 2017.

Massimo Massa, David Schumacher, and Yan Wang. Who is afraid of blackrock? The Review of Financial Studies, 34(4):1987–2044, 2021.

Daniel Neuhann and Michael Sockin. Financial market concentration and misallocation. Journal of Financial Economics, 159:103875, 2024.

Daniel Neuhann, Mahyar Sefidgaran, and Michael Sockin. Portfolio regulation of large financial institutions. Available at SSRN 3971043, 2021.

Dömötör Pálvölgyi and Gyuri Venter. Multiple equilibria in noisy rational expectations economies. Available at SSRN 2524105, 2015.

Joel Peress. Wealth, information acquisition, and portfolio choice. The Review of Financial Studies, 17(3):879–914, 2003.

Marzena Rostek and Marek Weretka. Price inference in small markets. Econometrica, 80(2): 687–711, 2012.

Marzena Rostek and Marek Weretka. Dynamic thin markets. The Review of Financial Studies, 28(10):2946–2992, 2015a.

Marzena Rostek and Marek Weretka. Information and strategic behavior. Journal of Economic Theory, 158:536–557, 2015b.

Aldo Rustichini, Mark A Satterthwaite, and Steven R Williams. Convergence to efficiency in a simple market with incomplete information. Econometrica: Journal of the Econometric Society, pages 1041–1063, 1994.

Mark A Satterthwaite and Steven R Williams. Bilateral trade with the sealed bid k-double auction: Existence and efficiency. Journal of Economic Theory, 48(1):107–133, 1989.

Jeff Tjornehoj. Exploring fund industry concentration: The good, the bad, and the unknown. Technical report, Broadridge Financial Solutions, 2018.

Hal R Varian. Microeconomic analysis. WW Norton & Company, 1992.

Dimitri Vayanos. Strategic trading and welfare in a dynamic market. The Review of Economic Studies, 66(2):219–254, 1999.

Dimitri Vayanos and Jean-Luc Vila. Equilibrium interest rate and liquidity premium with transaction costs. Economic theory, 13(3):509–539, 1999.

Dimitri Vayanos and Jiang Wang. Theories of liquidity. Foundations and Trends® in Finance, 6(4):221–317, 2012.

Xavier Vives. Aggregation of information in large cournot markets. Econometrica: Journal of the Econometric Society, pages 851–876, 1988.

Xavier Vives. Strategic supply function competition with private information. Econometrica, 79(6):1919–1966, 2011.

Robert Wilson. A bidding model of perfect competition. The Review of Economic Studies, 44 (3):511–518, 1977.

Robert Wilson. Auctions of shares. The Quarterly Journal of Economics, 93(4):675–689, 1979.

Kathy Yuan. Asymmetric price movements and borrowing constraints: A rational expectations equilibrium model of crises, contagion, and confusion. The Journal of Finance, 60(1):379–411, 2005.