Financial Market Concentration and Misallocation*

Daniel Neuhann† Michael Sockin‡

July 19, 2023

Abstract

How does financial market concentration affect capital allocation? We propose a complete-markets model in which real investment and financial price impact are jointly determined in general equilibrium. We identify a two-way feedback mechanism whereby price impact induces misallocation and misallocation raises price impact. The mechanism is stronger if productivity is low or productivity dispersion is high. Given rising dispersion, the model can rationalize trends in corporate discount rates, cash holdings, investment, asset prices, and capital reallocation over the last two decades, even when market concentration is relatively stable. Overall, our findings suggest that financial market concentration may hamper allocative efficiency.

*This paper supersedes a previous working paper by the same authors titled "Risk Sharing, Investment, and Asset Prices According to Cournot and Arrow Debreu." First version: January 2019. We thank Dimitris Papanikolaou (the editor), an anonymous referee, Ben Hebert, Gian Luca Clementi, Nuno Coimbra, Winston Dou, Wenxin Du, Selman Erol, Emmanuel Farhi, William Fuchs, Kinda Hachem, Zhengyang Jiang, Guido Lorenzoni, Yiming Ma, Andrey Malenko, Gregor Matvos, Aaron Pancost, Dimitris Papanikolaou, Robert Townsend, Lukas Schmid, Vish Viswanathan, Ariel Zetlin-Jones, and seminar participants at 2nd HKUST-Jinai Macroeconomics Workshop, NASMES 2019, SED 2019, EFA 2019, 10th Tepper-LAEF Advances in Macro-Finance conference, UBC Winter Finance Conference 2020, MFA 2020, AFA 2021, the Virtual Finance Theory Seminar, Boston College, Carnegie Mellon, and UT Austin for helpful comments. We thank Mahyar Sefidgaran for excellent research assistance.

†UT Austin. daniel.neuhann@mccombs.utexas.edu

‡UT Austin. michael.sockin@mccombs.utexas.edu

1 Introduction

A striking fact about the U.S. economy since the 2000s is that corporate investment has been weak despite falling risk-free rates and low costs of capital.[1] During this time, large nonfinancial firms also increased their net lending and accumulation of financial assets, such as corporate debt, in what has often been referred to as a "corporate savings glut" (e.g., Gruber and Kamin (2015)). Why were firms unwilling to invest despite low costs of capital? One popular explanation is a dearth in profitable investment opportunities. However, it is unclear which structural changes would have created such scarcity, and different explanations may have different policy and welfare implications.

In this paper, we propose one such alternative explanation: high financial market concentration and cross-sectional dispersion in investment opportunities have made it difficult for large firms to share risk and efficiently reallocate capital via financial markets. Our approach is motivated by the empirical fact that many financial markets, including those for corporate credit, have grown increasingly concentrated since the 1980s, with a large share of capital now managed by a relatively small number of large financial institutions and nonfinancial firms.[2]

While previous literature has shown that such concentration can distort asset prices and liquidity, our contribution is to study the feedback to real investment using a tractable general equilibrium model of strategic trading with rich heterogeneity in investment opportunities, scale effects in preferences, and nonlinear price impact. Our key theoretical result is a novel two-way feedback mechanism between financial price impact and capital misallocation, whereby price impact induces capital misallocation by impairing risk sharing and misallocation increases price impact by distorting the cross-sectional distribution

[1] Gutierrez and Philippon (2017), Fernald, Hall, Stock, and Watson (2017), and Alexander and Eberly (2018) find a decline in investment relative to trend since the early 2000s, and in particular relative to measures of Tobin's Q. Laubach and Williams (2016) and Del Negro, Giannone, Giannoni, and Tambalotti (2018) document a secular decline in risk-free rates over the past few decades; Bianchi, Lettau, and Ludvigson (2020) find a concurrent decline in risk premia.

[2] Corbae and Levine (2018) shows that the five largest U.S. banks held 47% of total U.S. bank assets in 2015; in the U.K., France, Germany, Italy, and Canada, the range is from 71% to 84%. The OCC estimates that over 90% of the notional amount of interest-rate swaps is accounted for by four banks, while over 95% of the CDS market is accounted for by three banks. In the corporate bond market, Li and Yu (2022) document that the median investment-grade bond is held by only 47 investors, while Celik, Demirtas, and Isaksson (2020) show 25 large nonfinancial firms alone held $356 billion in corporate bonds in 2018. Ben-David, Franzoni, Moussawi, and Sedunov (2021) show that the largest institutional investor oversaw 6.3% of total U.S. equity assets in 2016, while the top 10 investors managed 26.5%.

of cash flows. Based on this mechanism, we show that the distortions induced by market power are particularly severe when investment opportunities are highly dispersed across firms. Given the secular increase in productivity dispersion over the last 20 years documented by Cunningham, Foster, Grim, Haltiwanger, Pabilonia, Stewart, and Wolf (2022), we find that our model can rationalize many salient trends regarding corporate discount rates, the cost of capital, and investment behavior over this period. By exacerbating issues of liquidity, financial market concentration can therefore reduce the efficiency of capital allocation in a manner that aligns with recent empirical trends.

Our framework is a general equilibrium model with complete financial markets and a finite number of large investors that represent large firms. Investors allocate funds across risky individual-specific investment technologies, a safe storage opportunity with low returns (cash), and financial securities. Because firms in practice are known to engage in hedging, we assign to them risk-averse preferences that might reflect managerial risk aversion (e.g., Papanikolaou and Panousi (2012)), a desire to smooth cash flows, or large shareholder under-diversification (e.g., Greenwald and Stiglitz (1990)).[3] Because our model is quite general, one can also interpret very unproductive firms as financial institutions who lend and borrow from other firms.

Financial positions are unrestricted: firms can buy or sell any number of securities consistent with their budget constraints. The only friction in our model is imperfect competition in financial markets. In particular, we embed the canonical trade-off studied by the literature following Kyle (1989) that price impact deters trading and leads to unrealized gains from trade. Empirically, Koijen and Yogo (2019) and Bretscher, Schmid, Sen, and Sharma (2022) estimate that large investors have *high and persistent* price impact in corporate equity and bond markets, respectively, suggesting that price impact is relevant for capital reallocation.

From a theoretical perspective, the main novelty introduced by endogenous investment is that the level, riskiness, and cross-sectional distribution of cash flows are determined in general equilibrium alongside price impact. To achieve this in a tractable manner, we study a Cournot-Walras equilibrium with a competitive fringe composed of households and other price-taking investors, rather than an equilibrium in demand

[3] Amel-Zadeh, Kasperk, and Schmalz (2022), for instance, document that up to one-fifth of the largest U.S. firms has a nonfinancial blockholder or insider as its largest shareholder.

schedules among strategic investors only.[4] This allows us to incorporate not only scale effects and heterogeneity in investment opportunities, but also nonlinear price impact, all of which critically influence the feedback between investment and price impact.

In the model, financial markets allow firms to manage risk and reallocate capital. This dual role is consistent with practice. For example, Ferreira (2021) documents that Braeburn Capital, a wholly-owned subsidiary of Apple Inc., manages $244 billion in financial assets, representing 70% of Apple's total book assets, $153 billion of which was invested in corporate bonds. According to the *Wall Street Journal*, "Apple acts like a hedge fund by supporting this portfolio with $115 billion in debt." Apple is far from unique: corporate savings have risen in recent years, and large nonfinancial corporations are now net lenders rather than borrowers (e.g., Chen, Karabarbounis, and Neiman (2017)). Understanding how corporate trading affects capital allocation is therefore an urgent matter.

Our analysis offers three main results. The first is a characterization of an adverse feedback loop between price impact and capital misallocation. Price impact gives rise to capital misallocation through the established mechanism that it leads to unrealized gains from trade in financial markets. Our contribution is to show that the manner in which price impact affects investment differs across the productivity distribution. High-productivity firms are net borrowers in financial markets. For these firms, price impact hampers risk management and deters borrowing, and so they invest less than they would in competitive markets. Low-productivity firms instead are net suppliers of capital. For these firms, price impact deters lending to other firms, and so they invest *more* in their own investment opportunities than they would under perfect competition. As such, price impact induces capital misallocation. Moreover, because it also impedes risk sharing, investment in risky capital eventually falls, and firms increasingly self-insure through inefficient cash holdings. Capital misallocation then, in turn, exacerbates price impact in financial markets. This is because lower output growth among firms that stems from inefficient investment forces them to buy more / sell less assets to the competitive fringe. This lowers the fringe's consumption and, if the fringe has convex marginal utility (such as with constant relative risk aversion preferences), increases price impact. The amplification from this feedback loop is ultimately determined by the degree of market concentration and the cross-sectional distribution of agents' productivity, which are the

[4]The online appendix provides a detailed comparison of the two equilibrium concepts.

key fundamentals in the economy. Our analysis shows that ignoring the feedback to investment *understates* the consequences of market concentration for allocative efficiency.

Our second main result examines the general equilibrium relation between the distribution of cash flow risk in the economy and aggregate economic conditions. We find financial market distortions are sensitive to changes in fundamental gains from trade across market participants, such as shocks to the cross-sectional dispersion of investment opportunities. This is true even when such shocks would be entirely neutral in competitive markets. For example, a mean-preserving spread of *diversifiable* investment risk leads to higher price impact, and thus a bigger decline in risky investment. An increase in potential gains from trade therefore leads to fewer *realized* gains from trade. Because dispersion in firm productivity has increased over time (e.g., Cunningham, Foster, Grim, Haltiwanger, Pabilonia, Stewart, and Wolf (2022)), our analysis suggests a secular increase in the *distortions* from market power despite the absence of any *direct* changes in market concentration. Because dispersion is counter-cyclical, our model also suggests that liquidity is pro-cyclical, while reallocation is counter-cyclical. The former is consistent with evidence from NÆs, Skjeltorp, and Ødegaard (2011) for stock markets, and from Bao, Pan, and Wang (2011) and Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga (2021) for bond markets; the latter is documented by Eisfeldt and Ramipini (2006).

Our third main result pertains to asset pricing implications. In particular, we provide conditions under which all asset prices rise, the risk-free rate falls, and the market risk premium remains low as markets become concentrated. Although low investment amidst low costs of capital is typically seen as a puzzle, in our model they are a joint outcome of endogenous distortions to financial market trading. The relation between investment and asset prices is nonlinear. When market concentration is low, shocks to market concentration reallocate investment among firms and are primarily reflected in rising asset prices. As market concentration increases, firms increasingly self-insure using cash holdings. From this point on, incremental price adjustments are largely driven by changes in the quantity of risk. An interesting implication is that variation in market power may be difficult to detect using reduced-form measures of illiquidity, such as the price elasticity, because these also reflect changes in the quantity and distribution of risk.

To illustrate the empirical relevance of these channels, we use our model to interpret trends in corporate investment and financial returns from 2002 to 2016. Gormsen

and Huber (2022) document a striking fact from this period, which is that corporations report large wedges between their discount rates (or hurdle rates) and their estimates of the relevant weighted average cost of capital in financial markets. In particular, they show that the risk-free rate and the cost of capital have fallen sharply over this period, while discount rates have remained high and stable. Since private and market-based valuations do not align in the presence of price impact, our model predicts precisely such a wedge between hurdle rates and cost of capital. Calibrating our model to 2002, we show that an exogenous increase in productivity dispersion taken directly from Cunningham, Foster, Grim, Haltiwanger, Pabilonia, Stewart, and Wolf (2022) can account for most of the time variation in all three rates of return from 2002 to 2016 alongside anemic real investment. Adding a small increase in market concentration, consistent with data from Kwon, Ma, and Zimmermann (2023), leads to an even better fit with the cost of capital, and predicts a decline in investment and an increase in corporate cash holdings over the same period, all of which are consistent with the data. While we leave a full quantitative evaluation of our theory to future work, these findings suggest that financial market concentration may have contributed to the joint dynamics of real and financial variables over the last two decades. They also highlight one of our key theoretical results, which is that the distortions from financial market power are tightly linked to the cross-sectional distribution of investment opportunities.

The paper is organized as follows. Section 2 describes our model. Section 3 contains the theoretical analysis. In Section 4, we interpret recent trends in corporate investment relative to market rates of return through the lens of our model. Section 5 contains additional empirical predictions for understanding corporate hedging behavior and asset returns in a production-based framework. Section 6 concludes. Proofs are in Appendix A. Appendix B contains a description of our data.

1.1 Related literature.

Our paper contributes to a growing literature at the intersection of macroeconomics, financial markets, and industrial organization. Gabaix and Koijen (2020) posit institutional mandates and portfolio constraints limit large competitive investors' ability to absorb demand shocks in financial markets. Koijen and Yogo (2019), Haddad, Huebner, and Loualiche (2021), and Bretscher, Schmid, Sen, and Sharma (2022) use a demand-system

approach to analyze large investors behavior in financial markets, emphasizing the role of quantities for estimating the degree of price impact. However, they generally maintain the assumption of price-taking behavior and ignore feedback to real investment.

Other papers in this area focus on imperfect competition among financial intermediaries (e.g., Egan, Hortacsu, and Matvos (2017), Drechsler, Savov, and Schnabl (2017), Corbae and D'Erasmo (2021), Hachem and Song (2021)), some of which consider implications for investment. In contrast to these papers, we focus on imperfect competition among large firms that internalize their trades move asset prices. This allows us to link market concentration directly to trends in firm investment behavior and capital allocation. In this context, our focus on imperfect competition in financial markets distinguishes our work from studies of market concentration in product markets (e.g., Gutiérrez and Philippon (2017), Jones and Philippon (2016), Azar and Vives (2021), Corhay, Kung, and Schmid (2020), Chen, Dou, Guo, and Ji (2020)), which also lowers investment and risk-free rates but often raises productivity and risk premia through procyclical markups.

Our work is related to the literature on capital misallocation.[5] Eisfeldt and Ramippini (2006) provide evidence misallocation is counter-cyclical while reallocation is procyclical. Carlstrom and Fuerst (1997) and Ai, Li, and Yang (2020), for instance, relate misallocation to agency frictions, while Kurlat (2013) and Bigio (2015) link capital misallocation to asymmetric information about the endogenous quality of capital in secondary markets. We provide a novel channel through which strategic considerations induce large firms and financial institutions to voluntarily misallocate capital in illiquid financial markets, in which the endogenous degree of illiquidity is increasing in the extent of misallocation.

Our focus on risk management and capital allocation is related literature on corporate participation in financial markets. Ferreira (2021) shows nonfinancial firms have substantial holdings of corporate bonds and how these cross-holdings can account for the equity premium in a quantitative model. Ma (2019) provides evidence that firms act as strategic cross-market arbitrageurs in their own debt and equity securities. Closest to us are papers exploring how imperfect competition in product markets interacts with hedging in competitive forward markets (e.g., Allaz (1992), Allaz and Vila (1993), Cox and Karam (2022)). Imperfect hedging in those settings arises from the rich interaction between current hedging and future competition in *product markets.* We shut down prod-

[5]See Eisfeldt and Shi (2018) for a review of this literature.

uct market competition to study the economic consequences of the strategic rationing of securities in illiquid financial markets, with associated feedback effects to real investment.

Our paper contributes to the literature linking investment and asset markets in the tradition of Cochrane (1991) and Cochrane (1996). Kogan and Papanikolaou (2012) provide a review of this literature. One key insight from this approach is that a return on firm assets have a tight connection with the return on its equity. We show this connection is modified in particular ways when financial markets are illiquid, and find this has strong implications for investment. More recently, Schmid, David, and Zeke (2020) show heterogeneity in firm productivity leads to heterogeneity in firm-specific risk premia that can explain a sizable part of the observed dispersion in marginal products of capital. Our approach gives rise to heterogeneity in marginal products of capital as well as ex ante capital misallocation and a disconnect between stock returns and investment hurdle rates.

Our equilibrium concept is Cournot-Walras equilibrium in the tradition of Gabszewicz and Vial (1972). In this approach, strategic traders choose price-contingent quantities taking into account the quantities demanded by other strategic agents and the residual demand curve of the competitive fringe. A closely-related approach based on Kyle (1989) instead studies equilibrium in demand *schedules*. Rostek and Yoon (2020) provide a review of this literature, and papers such as Malamud and Rostek (2017) discuss how price impact hampers risk sharing given an exogenous endowment of securities or other endowments. Although this concept allows for a richer analysis of strategic interactions among large traders than Cournot-Walras, mainly by permitting agents to submit demand *schedules*, it typically requires strong assumptions on preferences and payoffs (such as the canonical CARA-normal setting) to preserve tractability. One contribution of our paper relative to this literature is that we introduce a real investment decision, which means that the distribution of cash flows (and therefore trading needs) are endogenously determined in general equilibrium alongside price impact. We provide a more detailed comparison of the two equilibrium concepts in the online appendix.

2 Model

In this section, we describe our model. There is a single good (the numeraire) and two dates, $t = 1, 2$. At date 1, agents make production and savings decisions under uncer-

tainty because the state of the world at date 2, z, is unknown. As a consequence, asset prices and production plans are jointly determined in equilibrium at $t = 1$. The set of possible states at date 2 is $\mathcal{Z} \equiv \{1, 2, \ldots Z\}$, and the probability of state $z \in \mathcal{Z}$ is $\pi(z) \in (0, 1)$ from the perspective of all agents. At $t = 2$, payoffs are realized and all agents consume.

There are two classes of agents: a discrete number of *strategic agents* who are large relative to the economy and internalize their price impact in financial markets, and a unit continuum of atomistic agents called the *competitive fringe* who take prices as given. Strategic agents represent large firms or large financial institutions who lend to firms. There are N *types* of strategic agents, indexed by $i \in \{1, 2 \ldots, N\}$, where a type indexes the production technology available to the agent. There are $1/\mu$ agents of every type, each of which has mass $\mu \in (0, 1]$. The total number of strategic agents is thus N/μ and the total mass is N. We use μ to vary the degree of market concentration without affecting the aggregate production possibility frontier. As $\mu \to 0$, we approach the perfect competition benchmark where there is an infinite number of infinitesimal agents. In what follows, we focus on the case in which all agents within a type follow symmetric strategies.

Strategic agent j of type i receives an initial endowment μe at date 1 and has access to a type-specific production technology that transforms $\mu k_{j,i}$ units of the numeraire at date 1 into $\mu y_i(z) k_{j,i}$ units of the numeraire in state z at date 2. Since the total endowment owned by agents of type i is e, parameter μ also determines the fraction of total initial wealth an agent of type i has relative to all agents of type i. That is, μ measures relative size. All endowments and technological payoffs are bounded, and production may be subject to agent-specific and aggregate risk. We place no other restrictions on $\{y_i(z)\}_{i=1}^N$ other than that total risky investment $\sum_{i,j} k_{i,j}$ is lower under financial autarky (i.e., when no financial trading is permitted) than under perfect competition.

Assumption 1 (Low Investment in Autarky) *Parameters are such that aggregate risky investment $\sum_{i,j} k_{i,j}$ is higher under perfect competition than when no financial trading is permitted.*

Strategic agents also have access to a risk-free storage technology that transforms $\mu s_i \geq 0$ units of the numeraire into $R\mu s_i$ units at date 2. Storage represents a safe alternative, such as cash, that is less productive on average, i.e., $E[y_i(z)] > R$ for all i. This flexible setting allows us to capture heterogeneity in the types of market participants. For example, an institutional investor may have a real investment opportunity that is nearly

equivalent to cash ($E[y_i(z)] \approx R$), whereas a technology firm may be highly productive ($E[y_i(z)] >> 0$).

Strategic agents have preferences over consumption at both dates that are represented by the utility index $u(c)$. Unless otherwise stated, $u(c)$ is homothetic, strictly increasing, strictly concave, and twice continuously differentiable. For simplicity, we do not impose non-negativity of consumption, but our results extend to this case. Assuming homothetic preferences allows us to highlight how firm size affects investment purely through equilibrium interactions rather than by directly affecting preferences. Risk aversion captures the notion that even large firms can exhibit risk aversion under a variety of frictions (e.g., Greenwald and Stiglitz (1990), Papanikolaou and Panousi (2012)).[6] Section 2.2 provides a further discussion of these assumptions.

In addition to the set of strategic agents, there is a non-strategic competitive fringe with mass m_f composed of price-taking agents that represent households or other competitive traders. In contrast to strategic agents, the fringe has linear preferences over consumption at date 1, and is risk-averse over consumption at date 2 with von Neumann-Morgenstern utility index, $u_f(c)$. We assume $u_f(c)$ is strictly increasing, strictly concave, and twice continuously differentiable. The competitive fringe receives endowment of the numeraire e at date 1 and $e_{2,f}(z)$ at date 2 in state z. The fringe has no production technology, reflecting its role as households and retail investors in financial markets. Quasi-linear preferences permit a particularly tractable demand system, but the main economic forces are unchanged if the fringe were also risk averse over consumption at date 1.

Both strategic agents and the competitive fringe participate in financial markets at date 1. The set of available assets is the complete set of Arrow securities. That is, there are $|\mathcal{Z}|$ securities, and security z pays one unit of the numeraire at date 2 in state z but zero otherwise. This ensures that markets are complete from a spanning perspective. We discuss the case of redundant securities in Subsection 3.1.

We denote by $a_{j,i}(z)$ the asset holdings of the state z security by agent j of type i, where $a_{j,i}(z) < 0$ denotes a sale. The aggregate position of type i strategic agents and all

[6] Asplund (2002) and De Giovanni and Iakimova (2022) also model strategic, risk-averse firms but focus on product market competition rather than imperfect competition in financial markets. Papanikolaou and Panousi (2012) not only provide a theory of firm risk aversion based on manager preferences, but also provide evidence of this channel in the set of public firms in Compustat.

strategic agents are, respectively,

$$a_i(z) \equiv \sum_{j=1}^{1/\mu} \mu a_{j,i}(z) \qquad \text{and} \qquad A(z) \equiv \sum_{i=1}^{N} a_i(z).$$

The competitive fringe's position in security z is $a_f(z)$. Market clearing conditions are

$$A(z) + m_f a_f(z) = 0 \qquad \text{for all } z. \tag{1}$$

We define $\mathbf{A}$ to be the $(N+1) \times Z$ matrix of asset holdings for all agents and all assets, with entries $\sum_j^{1/\mu} \mu a_{j,i}(z)$ that sum asset demands across the J_i agents within each type. The row cardinality of $\mathbf{A}$ is $N+1$ because there are N types strategic agents and a competitive fringe. The market-clearing price of asset z given $\mathbf{A}$ is then denoted by $Q(\mathbf{A}, z)$.

2.1 Decision Problems and Equilibrium Concept

Our equilibrium concept is Cournot-Walras in the tradition of Gabszewicz and Vial (1972). In this approach, strategic traders place price-contingent orders taking into account the quantities demanded by other strategic agents and the residual demand curve of the competitive fringe. As discussed at the outset, this concept simplifies strategic interactions among traders relative to the equilibrium concept in the tradition of Kyle (1989), but it allows us to incorporate rich heterogeneity in agent productivity, asymmetric equilibria, and nonlinear residual demand in a tractable manner, all of which are important for our insight that there is a two-way feedback between price impact and real misallocation. In line with Gabaix and Koijen (2020), the limited risk-bearing capacity of the fringe can also capture that there are few traders who quickly reallocate capital in response to shocks.

A *strategy* σ_f for the competitive fringe consists of asset positions and consumption, $\sigma_f = \{\{a_f(z)\}_{z \in \mathcal{Z}}, c_{1,f}, c_{2,f}\}$. Since the competitive fringe takes prices as given, its perceived pricing function in each state a constant, $\tilde{Q}_f(\mathbf{A}, z) = \tilde{q}(z)$ for some function

$\tilde{q}(z)$. The fringe's decision problem is

$$\begin{aligned} U_f = \max_{\sigma_f} \quad & c_{1f} + \sum_z \pi(z) u_f(c_{2,f}(z)) \\ \text{s.t.} \quad & c_{1f} = e - \sum_z \tilde{q}_f(z) a_f(z), \\ & c_{2,f}(z) = e_{2,f}(z) + a_f(z). \end{aligned} \tag{2}$$

A strategy $\sigma_{j,i}$ for strategic agent j of type i consists of asset positions, investments, and consumption, $\sigma_{j,i} = \{\{a_{j,i}(z)\}_{z\in\mathcal{Z}}, s_{j,i}, k_{j,i}, c_{1,j,i}, c_{2,j,i}(z)\}$. When deciding on an optimal strategy, strategic agents must form beliefs over the residual inverse demand function that maps aggregate asset portfolios into prices, *given* the asset positions of all other agents as summarized by the vector of other agents' strategies $\sigma_{-j,i}$. We denote the *perceived pricing function* used by agent j of type i to forecast her influence on the price of security z by $\tilde{Q}_{i,j}(\mathbf{A}, z)$. The associated decision problem is

$$\begin{aligned} U_{j,i} = \max_{\sigma_{j,i}} \quad & u\left(c_{1,j,i}\right) + \sum_{z\in\mathcal{Z}} \pi(z)\, u\left(c_{2,j,i}(z)\right) \\ \text{s.t.} \quad & \mu c_{1,j,i} = \mu e - \mu k_{j,i} - \mu s_{j,i} - \mu \sum_{z\in\mathcal{Z}} \tilde{Q}_{i,j}(\mathbf{A}, z) a_{j,i}(z), \\ & \mu c_{2,j,i}(z) = \mu y_i(z) k_{j,i} + \mu a_{j,i}(z) + \mu R s_{j,i}. \end{aligned} \tag{3}$$

We define preferences and controls in this manner recognizing that the consumption of strategic agent j of type i is actually $\mu c_{1,j,i}$ and $\mu c_{2,j,i}(z)$ at dates 1 and 2, respectively, and similarly with optimal asset holdings and investment. Given homothetic utility, however, optimal policies are invariant to μ.

Definition 1 *A Cournot-Walras equilibrium consists of a strategy $\sigma_{j,i}$ for each strategic agent, a strategy σ_f for the competitive fringe, and pricing functions $Q(\mathbf{A}, z)$ for all $z \in \mathcal{Z}$ such that:*

1. *Fringe optimization: σ_f solves decision problem (2) given $\{\tilde{q}_f(z)\}_{z\in\mathcal{Z}}$*
2. *Strategic agent optimization: For each agent j of type i, $\sigma_{j,i}$ solves decision problem (3) given (i) other agents' strategies $\{\sigma_{-j,i}, \sigma_f\}$ and perceived pricing functions $\{\tilde{Q}_{j,i}(\mathbf{A}, z)\}_{z\in\mathcal{Z}}$.*
3. *Market-clearing: Each market clears with zero excess demand according to (1).*
4. *Consistency: all agents have rational expectations, which requires for strategic agents that $\tilde{Q}_{j,i}(\mathbf{A}, z) = Q(\mathbf{A}, z)$ for all i, j and z.*

Similar to product market models of Cournot competition, strategic interaction in financial markets is intermediated by a group of price-taking agents, i.e., the competitive fringe. While a strategic agent takes the asset positions of other strategic agents as given, he does internalize how his own demand impacts equilibrium asset prices by altering the marginal utility of the fringe. Through this channel, how one strategic agent type trades affects how another strategic agent trades by altering the price and price impact that agent faces. When he increases his demand, asset prices and price impact increase. This reduces the demand of other strategic agents and worsens strategic distortions. Although the utility function and size (i.e., m_f) of the fringe does affect the shape of the equilibrium price function, it does not change the manner in which strategic agents fundamentally impact each other's behavior.

2.2 Model Discussion

We now briefly discuss several of our modeling assumptions. First, we model the objective of strategic agents as expected utility maximization. This is equivalent to the more standard approach of shareholder value maximization for the case of a single large private shareholder, and can be easily extended to the case of heterogeneous large shareholders with fixed holdings across agents (i.e., common ownership) provided we specify how voting rights are allocated. A subtlety of our analysis is that although financial markets are complete, such large shareholders would not necessarily agree on production plans because the strategic agents in which they invest do not share risks efficiently in equilibrium. This renders the problem similar to those studied in the intractable "stock market" equilibria of Radner (1974) and Dréze (1974).[7] We opt for this parsimonious objective to avoid these issues and focus on the unique implications of our channel of how strategic trading in financial markets interacts with firm behavior.

Second, we consider a model with only two dates. One might worry with multiple periods that imperfect risk sharing would be irrelevant for production decisions because agents can self-insure with cash or credit lines in lieu of trading in financial markets (e.g., Bolton, Chen, and Wang (2011)). This is not the case for two reasons. First, the return to

[7] Issues of the objective of the firm with heterogeneous shareholders are not specific to our mechanism of imperfect competition in financial markets, and are beyond the scope of our paper. How a firm chooses its marginal valuation of production across states (i.e., its shareholder-aggregated state prices) is irrelevant to our insight that the firm will choose its financial asset positions to put a wedge between these private valuations and market prices.

a risk-free portfolio in financial markets is bounded from below by that on storage R and would be from above by the competitive lending rate on any private credit line. As such, financial market participation is preferable to self-insurance unless market concentration distortions are sufficiently severe, which our model captures. Second, firms under-insure regardless of initial wealth. As shown in Neuhann and Sockin (2023) in the context of forward-looking investors, such investors become even more exposed to their own income shocks over time, amplifying their strategic under-diversification. The distortions we characterize will therefore remain relevant in a dynamic setting.

3 Equilibrium

We now characterize the equilibrium. In models of strategic interaction, a key object is the equilibrium functional that determines prices and price impact. The first step is to derive this object in closed form. This then allows us to prove the existence of equilibrium and discuss the distortions induced by imperfect competition in financial markets.

3.1 Equilibrium Demand System and Price Impact

We derive the equilibrium pricing functional using the decision problem of the competitive fringe. Since the competitive fringe takes prices as given, the first-order conditions for portfolio optimality require that asset prices are equal to the fringe's marginal utility. This delivers an analytic solution for the pricing functional and price impact. This demand system has many useful properties. In particular, there is no arbitrage and the equilibrium is invariant to the introduction of redundant securities.

Proposition 1 (Demand System and Law of One Price) *The Law of One Price holds. All available assets are traded, but investment, consumption, and prices are invariant to the introduction of redundant assets. Arrow security prices are given by:*

$$Q(\mathbf{A}, z) = q\left(z\right) \equiv \pi(z) u_f'\left(c_{2,f}(z)\right) \qquad \textit{where} \qquad c_{2,f}(z) = e_{2,f}(z) - \frac{1}{m_f} A(z). \tag{4}$$

Price impact of strategic agent i is symmetric across agents and satisfies

$$\frac{\partial \tilde{Q}_{j,i}(\mathbf{A}, z)}{\partial a_i(z)} = \frac{\mu}{m_f} q'(z) \qquad \textit{where} \qquad q'(z) \equiv \frac{\partial q(z)}{\partial A(z)} = -\pi(z) u_f''\left(c_{2,f}(z)\right) > 0. \tag{5}$$

Because the competitive fringe takes price as given, its first-order condition yields a residual demand curve for every Arrow security.[8] Strategic agents optimize against this residual demand curve, taking as given the quantities demanded by other strategic agents. Large agents' portfolios thus pin down the *level* of fringe consumption, while price impact reflects the degree to which a marginal change in quantities affects fringe marginal utility at that level of consumption. Since marginal utility is nonlinear under standard preferences, strategic interactions among large traders influence both the level and slope of prices, giving rise to a price impact function. This extends the central insight from the literature following Kyle (1989) to an environment with endogenous cash flows, rich heterogeneity, wealth effects, and non-linear price impact.

Invariance with respect to redundant securities arises because any combination of assets that delivers the same consumption process to the competitive fringe induces the same prices and price impact. The role of μ is to scale each strategic agents' influence on the consumption of the competitive fringe. When $\mu \to 0$, this influence is negligible and price impact disappears. We are then back to the competitive benchmark. Finally, the price impact function is unique because it is fully pinned down by fringe marginal utility. This nullifies any strategic uncertainty that would give rise to equilibria multiplicity through self-fulfilling coordination on different price impact functions.[9]

That price impact depends on the fringe's consumption level also creates a link between aggregate output and strategic considerations. In particular, variations in output affects price impact as long as marginal utility varies with consumption. This lead to one direction of the feedback loop between real investment and financial markets, which is that declines in output due to misallocation or productivity shocks can boost price impact.

Corollary 1 (Real Allocations and Price Impact) *Define aggregate output in state z to be* $Y(z) = \sum_{i=1}^{N} y_i(z)k_i + Rs_i + m_f e_f(z)$*, and let fringe preferences satisfy convex marginal utility (such as CRRA). Holding investment choices fixed, a decline in aggregate output in state z because of lower productivity* $\{y_i(z)\}_{i=1}^{N}$ *and/or less risky investment* k_i *leads to higher price impact in state z.*

In complete markets, a decline in output must lead to a marginal decrease in consumption for all agents in the economy. Since price impact co-moves positively with fringe marginal

[8] Quasi-linearity ensures this residual demand curve is state-specific, but plays no substantive role.

[9] Kyle (1989) shows in the special case of the CARA-normal setting that the unique residual demand curve is linear. More generally, there may be many residual demand curves that can support an equilibrium.

utility when marginal utility is convex, shocks to productivity or to the efficient allocation of capital can raise price impact and the incentives for strategic distortions.

3.2 Optimal Strategic Portfolios and Investment Wedges

We now study the optimal policies of strategic agents, taking as given the demand system derived above. Without loss of generality, we assume all strategic agents of type i behave symmetrically. We define the *state price* $\Lambda_i(z)$ for state z and an agent of type i as the ratio of expected marginal utility in state z and marginal utility at date 1:

$$\Lambda_i(z) \equiv \frac{\pi(z)\, u'(c_{2,i}(z))}{u'(c_{1,i})}. \tag{6}$$

We can then characterize optimal portfolios and investment policies as follows.

Lemma 1 (Equilibrium Existence and Optimal Strategies) *There exists an equilibrium in which the optimal policies of agents of type i for $a_i(z)$, k_i and s_i, are homogeneous of degree 1 in e conditional on asset prices $q(z)$. These policies satisfy the optimality conditions*

$$\begin{aligned} a_i(z): &\qquad \Lambda_i(z) = q(z) + \frac{\mu}{m_f} q'(z) a_i(z), \\ k_i: &\qquad \sum_{z\in\mathcal{Z}} \Lambda_i(z)\, y_i(z) \leq 1 \ \ (\textit{and} = \textit{if } k_i > 0), \\ s_i: &\qquad \sum_{z\in\mathcal{Z}} \Lambda_i(z) \leq 1, \ \ (\textit{and} = \textit{if } s_i > 0). \end{aligned}$$

Conditional on state prices, the optimality conditions for risky investment and storage are standard: agents equate the state price-weighted expected return to the marginal cost of investing. Moreover, agents always invest some amount of capital in risky capital because it has higher returns than storage on average. In equilibrium, however, investment policies are distorted because price impact distorts state prices.

This can be seen in the first-order conditions for optimal portfolios. Rather than aligning state prices with market prices, optimal portfolios are shaped by endogenous *wedges* that appear because agents voluntarily misalign their state prices to tilt asset prices in their favor. Since buyers of a particular Arrow security reduce demand to lower prices, and sellers of the same Arrow security reduce supply to raise prices, state-specific wedges

$w_i(z)$ are negative for sellers and positive for buyers,

$$w_i(z) \equiv \frac{\mu}{m_f} q'(z) a_i(z). \tag{7}$$

One component of the investment wedge is the degree of price impact: if price impact increases, so must the wedge. This creates the opposite direction of the feedback loop from Corollary 1, which showed that changes in allocative efficiency can increase price impact. (For simplicity, we engineer an increase in price impact simply by varying the endowment of the competitive fringe. But we could also change fringe preferences, or the degree of market concentration.)

Corollary 2 (Price Impact and Investment Wedges) *Suppose the competitive fringe has convex marginal utility and we reduce its endowment in state z, $e_f(z)$. Then, both prices $q(z)$ and price impact $q'(z)$ (weakly) increase. As such, investment wedges must increase.*

The other component of the wedge is the trading volume $a_i(z)$, which determines the inframarginal benefit of a price change. This mechanism relates the market power friction to fundamental trading needs, such as idiosyncratic risk exposures that can be offset through financial markets. Under perfect competition, any increase in idiosyncratic dispersion is absorbed through higher trading volumes without affecting allocations or prices. With strategic agents, instead, an increase in gains from trade also increases incentives for rent-seeking behavior, and agents respond by distorting trading volumes more and realizing fewer gains from trade. We illustrate this insight in Corollary 3.

Corollary 3 (Cross-Sectional Dispersion and Distortions) *Fixing an investment policy, state price dispersion and financial market wedges are increasing in mean-preserving spreads of $\{y_i(z)\}$.*

An important implication of this result is that, even absent *changes* in market concentration, the equilibrium *consequences* of market concentration may grow more severe if the need to share risk and reallocate capital increases. Empirical evidence indicates that idiosyncratic dispersion has indeed been increasing over the past two decades. In Section 4, we exploit this fact to evaluate trends in corporate investment through the lens of our model. Dispersion is also known to increase during recessions. According to our model, both strategic distortions *and* the benefits from trading are consequently counter-cyclical. Eisfeldt and Ramipini (2006) document that capital misallocation is pro-cyclical while the benefits of reallocation are counter-cyclical.

The previous results show that price impact distorts investment policies. However, it is not yet clear whether this induces over- or under-investment at the firm-level and in the aggregate. In this regard, a useful feature of our complete-markets setting is that the decision to buy or sell a particular security can be directly linked to an agent's production technology. An agent sells if she has high income in a particular state of the world, and buys if she has low income. Financial market distortions thus lead to a specific form of state price distortions: relative to the competitive benchmark, the marginal value of income goes *down* in states with high output and *up* in states with low output.

Lemma 2 (State price expansion) *Fixing investment decisions, price impact leads to a negative wedge when the agent is a seller and a positive wedge when the agent is a buyer. Price impact therefore leads to a decline in an agent's state prices in all states where the agent has above average output, and an increase in all states where the agent has below average output.*

The overall effect of price impact on optimal investment decisions is then apparent if we use the first-order conditions for optimal trading to substitute state prices. Optimality conditions for risky investment and storage, respectively, can be expressed as:

$$\sum_z q(z)y_i(z) + \sum_z w_i(z)y_i(z) = 1 \qquad \text{and} \qquad \sum_z q(z)R + \sum_z w_i(z)R \leq 1. \tag{8}$$

Similar to standard q-theory, an agent optimally invests to the point where the marginal private benefit is equal to the marginal cost (which is equal to one here.) However, price impact forces a wedge in valuations that is sensitive to asset prices and price impact. Agents for whom the *net wedge* $\sum_z w_i(z)y_i(z)$ is negative *under-invest* based on market prices $q(z)$ (i.e., $\sum_z q(z)y_i(z) > 1$) because they retain too much of their own production risk. Similarly, agents for whom the net wedge is positive *over-invest* because they buy too few securities from financial markets.

An important implication of this result is that, when markets are concentrated, market prices are not appropriate measures of private investment incentives. In particular, our model predicts a disconnect between firm investment hurdle rates and the return to capital measured at market prices. Such wedges are observed in practice (e.g., Jagannathan, Matsa, Meier, and Tarhan (2015)). Most recently, Gormsen and Huber (2022) provide evidence of a wedge in discount rates and financial market returns that acts as a drag on investment, consistent with the "factorless" income documented in Karabar-

bounis and Neiman (2018). This can explain not only low investment by a firm with a high Tobin's q, but also an increasing share of its profits accruing to shareholders. Most strikingly, Gormsen and Huber (2022) show that wedges are systematically larger when risk-free rates are lower. We show below this is precisely what our model predicts.[10] In Section 4, we use moments from Gormsen and Huber (2022) to discipline an empirical exercise using our model.

3.3 Equilibrium Investment and Misallocation

We now map wedges in the optimality conditions for investment into implications for the equilibrium allocation of capital in the cross-section of investment opportunities. In particular, we show that highly productive agents tend to under-invest when financial markets are concentrated, while unproductive agents tend to over-invest. These results depend critically on the fact that our model permits rich cross-sectional heterogeneity in investment opportunities, which is a novel contribution for models with price impact.

Formally, we identify high- and low-productivity agents by comparing their privately-optimal investment scales under two benchmarks: the efficient benchmark with perfect competition (zero price impact), and financial autarky (no financial markets). A high-productivity agent is one who invests more in the competitive benchmark than in autarky; a low-productivity agent is one who invests relatively more under autarky. Underlying this idea is the notion that the production technologies of agents who invest less under perfect competition than under autarky must be dominated by those of other agents.

Proposition 2 shows that market concentration distorts an agent's investment policies in a manner that depends on individual productivity levels. A high-productivity agent chooses a lower scale of production than under perfect competition. This is because it would be efficient for her to borrow from and share risks with low-productivity agents, but price impact hampers the realization of these gains from trade. A low-productivity agent, in contrast, chooses a higher scale of production than under perfect competition as price impact deters the reallocation of capital to the more productive agents. When market concentration is high, agents also increasingly rely on inefficient storage to self insure. Through these mechanisms, market concentration lowers aggregate investment

[10] Pushing the interpretation further, our mechanism offers an explanation for the declining labor share of profits in recent decades, i.e., the labor share puzzle. It is particularly compelling because we abstract from product market competition, which Rognlie (2018) shows cannot explain the labor share puzzle.

and induces misallocation.

Proposition 2 (Equilibrium Investment) *As a result of market concentration μ:*

(i) the optimal scale of risky production for high-productivity firms is lower,

(ii) the optimal scale of risky production for low-productivity firms is higher,

(iii) for μ sufficiently large, aggregate risky investment and average productivity are lower,

(iv) corporate cash holdings (investment in safe storage) are (weakly) higher.

3.4 Two Canonical Examples

We now use two canonical examples to illustrate how strategic considerations alter investment decisions and feed back into price impact. In the first, we analyze ex-ante symmetric agents with purely diversifiable risk to show how strategic interactions in financial markets reduce risky investment. In the second, we examine two agents when one has a dominated technology that would not be utilized in the competitive equilibrium to show how they can lead to cross-sectional misallocation. For simplicity, in these two examples we focus on a limit in which the size of the competitive fringe, m_f, is arbitrarily small (i.e., $m_f \to 0$), while each strategic agent's relative size is fixed, (i.e., $\frac{\mu}{m_f} = \kappa$).[11]

We first consider two equally productive types facing pure idiosyncratic risk. The efficient outcome in this case is that agents invest in their risky technologies and then fully share risks via financial markets. The key effect of financial market concentration is to distort risk sharing, which then feeds back into inefficiently low investment.

[11]The limit in which the fringe becomes arbitrarily small (i.e., $m_f \to 0$) is a specific equilibrium in which asset prices are equal to the average state prices across the strategic agents

$$q(z) = \frac{1}{N} \sum_{z \in \mathcal{Z}} \Lambda_i(z) \ \forall z. \tag{9}$$

Price impact is given by $\kappa q'(z)$, and still measures how a marginal change in the demand of a strategic agent affects the asset price $q(z)$ by altering the competitive fringe's marginal utility. Although the fringe absorbs a trivial amount of the strategic agents' net asset demand when $m_f = 0$, i.e., $A(z) = 0$, the ratio $A(z)/m_f$ can converge to a nontrivial limit so the relation

$$q'(z) = \lim_{m_f \to 0} u_f'' \left(e_f(z) + \frac{A(z)}{m_f} \right), \tag{10}$$

holds. Consequently, the strategic interactions of Cournot competition are preserved when $m_f \to 0$, with the fringe providing the equilibrium price (impact) function but not buying or selling any asset.

Example 1 (Ex-ante Symmetry and Diversifiable Risk) *There are two equally likely states at date 2, $z \in \{1,2\}$ with $\pi(z) = \frac{1}{2}$, and two ex-ante symmetric types of strategic agents, $i \in \{1,2\}$. All agents have an initial endowment $\bar{y}$. Strategic agents face diversifiable production risk: $y_i(i) = \bar{y} + \Delta$ and $y_i(-i) = \bar{y} - \Delta$. That is, in either state one type has a high return and the other has a low return. The fringe, in contrast, receives $\bar{y}$ in every state.*

Since strategic agents are ex-ante symmetric, search for an equilibrium where each agent sells a_S units of the claim on the state in which she has high income, and buys a_B units of the claim on the other state. By market clearing in the limit where the fringe is small, we have $a_S = -a^*$ and $a_B = a^*$ for some a^*. All states must therefore have the same prices q^* and price impact q'^*. Moreover, net expenditures on financial claims at date 1 is 0. As such, a^* must satisfy

$$\textit{Seller optimality:} \qquad \frac{\frac{1}{2}u'((\bar{y}+\Delta)\,k - a^*)}{u'(\bar{y}-k)} = q - \kappa q' a^*.$$

$$\textit{Buyer optimality:} \qquad \frac{\frac{1}{2}u'((\bar{y}-\Delta)\,k + a^*)}{u'(\bar{y}-k)} = q + \kappa q' a^*.$$

Under perfect competition, $a^* = \Delta k$. As such, $0 < a^* < \Delta k$ with market concentration. Aggregating the wedges across states of the world gives the following net wedge:

$$-(\bar{y}+\Delta)\kappa q' a^* + (\bar{y}-\Delta)\kappa q' a^* = -2\Delta\kappa q' a^* < 0.$$

That is, the net wedge is negative for both types. As such, both agents will opt to invest less than under perfect competition because they cannot efficiently share production risks.

Two further implications follow. First, as long as Δ is large enough, a sufficiently large increase in μ leads agents to start self-insuring using cash. Hence, market concentration can lead to an increase in corporate savings. Second, because there is less output in every state when agents shift away from the (efficient) risky technology, prices and price impact must rise. Hence there is a two-way feedback between real allocations and price impact.

Next, we turn to a setting with scope for misallocation. The efficient allocation is that only Type 1's technology is used, but in equilibrium both types may invest.

Example 2 (Asymmetry with dominated technologies) *There are two types of strategic agents, $i \in \{1,2\}$. Production technologies satisfy $y_1(z) = y^h$ and $y_2(z) = y^l \in (R_f, y_h)$ so that Type 2's production technology is strictly dominated. All agents have an initial endowment $\bar{y}$. The fringe receives $\bar{y}$ in every state. Preferences are of the CRRA type.*

Since there is no risk, we can search for an equilibrium where agent 1 sells a_1 units of the claim to its production and agent 2 buys a_2 units. By market clearing in the limit where the fringe is small, we have $a_1 = -a_2 = -a^*$ for some a^*. This claim has price q^* and price impact q'^*. Moreover, strategic agents net expenditures on assets at date 1 are $k_1 - qa^*$ and $k_2 + qa^*$, respectively. As such, a^* must satisfy

$$\text{Seller optimality:} \qquad \frac{u'(y^h k_1 - a^*)}{u'(\bar{y} + qa^* - k_1)} = q - \kappa q' a^*.$$

$$\text{Buyer optimality:} \qquad \frac{u'(y^l k_2 + a^*)}{u'(\bar{y} - qa^* - k_2)} = q + \kappa q' a^*.$$

Under perfect competition, $k_2 = 0$ and $a^* > 0$. As such, $a^* > 0$ with market concentration. The wedges gives the following net wedges:

$$\text{Type 1}: \qquad -y^h \kappa q' a^* < 0.$$

$$\text{Type 2}: \qquad y^l \kappa q' a^* > 0.$$

Since the wedge is negative for Type 1 agents and positive for Type 2 agents, the agent with the dominated technology over-invests relative to the competitive benchmark. Since this reduces output, prices and price impact must rise. As such, misallocation feeds back into price impact.

3.5 Asset Pricing Implications

Financial market distortions are central to our theory of investment. Hence it is important to assess the model's asset pricing implications. Taking a cross-sectional average of the first-order condition for $a_i(z)$ yields the following expression for the price of the Arrow security for state z:

$$q(z) = E^*[\Lambda_i(z)] - \overline{mkt}\,(z)\,, \tag{11}$$

where $E^*[\cdot]$ is the cross-sectional average and $\overline{mkt}\,(z) = q'(z)\frac{A(z)}{N}$ the *average market concentration* in state z. The state price- and market-implied risk-free rates, r_f^* and r_f^m, are[12]

$$r_f^* = \left[\sum_{z\in\mathcal{Z}} E^*\left[\Lambda_i\,(z)\right]\right]^{-1} \qquad \text{and} \qquad r_f^m = \left[\sum_{z\in\mathcal{Z}} q(z)\right]^{-1}$$

[12]The wedge between r_f^* and r_f^m is determined by market concentration: $\frac{1}{r_f^*} - \frac{1}{r_f^m} = \sum_z \overline{mkt}\,(z)$.

The market risk premium is the expected excess return on the tradable portfolio that owns all risky production:

$$RP_{mkt} = \frac{E\left[\sum_{i=1}^{N} y_i(z)\, k_i\right]}{\sum_{z\in\mathcal{Z}} q(z) \sum_{i=1}^{N} y_i(z)\, k_i} - r_m \tag{12}$$

Lemma 3 shows that the market risk premium can be decomposed into two pieces. The numerator of the right-hand side of (13) is the classical risk premium based on the covariance of the market portfolio with the state prices of the competitive fringe. The denominator represents the total distortion to the marginal value of capital from strategic trading with price impact. Firms that under-invest because of limited corporate hedging have an inflated Tobin's q, while those that over-invest have a depressed Tobin's q. This distorts the cost of the market portfolio and consequently the market risk premium.

Lemma 3 (Market Risk Premium) *The market risk premium, RP_{mkt}, satisfies:*

$$RP_{mkt} = -\frac{Cov\left(\frac{q(z)}{\sum_{z\in\mathcal{Z}} q(z)}, \sum_{i=1}^{N} y_i(z)\, k_i\right)}{\sum_{i=1}^{N} k_i - \mu \sum_{z\in\mathcal{Z}} q'(z) \sum_{i=1}^{N} a_i(z)\, y_i(z)\, k_i}. \tag{13}$$

Having defined key asset pricing objects, we now turn to the effects of an increase in market concentration. To isolate the pure effects of financial market concentration on rates of return, in the following we assume the fringe is *passive* in the counterfactual competitive equilibrium, by which we mean that the fringe holds a non-negative position in every asset if $\mu = 0$. (There always exists a fringe endowment process such that this condition is satisfied.) This assumption is not necessary but convenient because it ensures that price differences between the market and the competitive counterfactual equilibrium are primarily governed by the interactions of strategic agents. It also rules out the case of a quasi-monopsony or monopoly in which all strategic agents trade in the same direction.

Assumption 2 (Passive Fringe) *$\{e_{2,f}(z)\}_{z\in\mathcal{Z}}$ is such that $a_f(z) \geq 0$ when $\mu = 0$.*

Our key asset pricing result for this section is a sharp characterization of the comparative statics of returns in the benchmark case where there is small number of large investors who are relatively similar ex-ante (but not ex-post). Analogous results obtain in the "strategic limit" where the mass of the fringe is small. This is the case we exploit in our numerical analysis below.

Definition 2 (Type-Symmetric) *Two agent types are type-symmetric if they have ex-ante symmetric income risks so that they face identical decision problems.*

Proposition 3 then establishes that market concentration raises all asset prices, depresses the risk-free, and simultaneously lowers the market risk premium.

Proposition 3 (Asset Prices and Risk-free Rate) *Suppose all strategic agents types are symmetric and that Assumption 2 holds. Then, as a result of market concentration μ:*

(i) If agents do not invest in storage, asset prices are higher state-by-state than in the competitive equilibrium. If agents employ storage, then some states may instead have lower prices.

(ii) The risk-free rates, (state-price implied) r_f^* and (market-implied) r_f^m, are lower than in the competitive equilibrium and bounded below by the rate of return on storage R.

(iii) The market risk premium is lower than in the competitive equilibrium.

Market concentration lowers the risk-free rate primarily because imperfect risk sharing raises the value of insurance. It lowers the market risk premium through both a quantity and a price of risk channel. The quantity of risk channel is straightforward. When investment misallocation is severe, such as when storage is employed, then there is less total risky production in the economy, and the quantity of risk falls. This lowers the market risk premium. The price channel is more subtle. For fixed investment policies, market concentration distorts trading in financial markets, and if the distortions to state prices favors sellers, this inflates security prices and reduces their correlation with total production. Both effects reduce the price of risk.

The most interesting implication is that market concentration can lead to a *joint* decline in investment, risk-free rates and the risk premium, a trend that has been observed in the U.S. over the past few decades. This combination is a priori surprising because low rates of return should lower costs of capital, leading to an increase in risky investment. In our setting, however, all three effects are symptoms of the same underlying cause, which is that strategic distortions in asset markets lead to misallocation and hamper risk sharing. Our result that strategic incentives are greater in periods of greater cross-sectional dispersion (such as recessions) further amplifies this mechanism.

Away from the symmetric benchmark, our mechanism is co-mingled with the well-known result that asset prices more closely reflect the preferences and/or strategic decisions of the largest agents. For instance, a single "monopolist" would push prices up for

all securities he chooses to sell and push them down for all securities he buys. The numerical exercise in the following section considers a setting with a number of asymmetric traders, and finds that the risk-free rate continues to decline in market concentration.

4 Recent Trends Through the Lens of the Model

A prominent literature (e.g., Gutiérrez and Philippon (2017), Crouzet and Eberly (2021)) documents that starting in the early 2000s, there has been a sharp decline in corporate investment and productivity when compared to previous trends and, more specifically, relative to financial measures of the cost of capital and investment opportunities (i.e, Tobin's Q) over the same period. One particular manifestation of this is the observation in Gormsen and Huber (2022) that firms' self-reported discount rates (or hurdle rates) far exceed their perceptions of their weighted average cost of capital, and that the wedge between the two has increased. In particular, discount rates have been relatively stable since 2002 despite sizable declines in the risk-free rate and the cost of capital. (See Figures 3 and 4 in Appendix B, which also includes a detailed description of the data).

Our theory suggests that market concentration can generate such a wedge between discount rates (i.e., firm-level marginal valuations of investment) and market-based costs of capital (i.e., the return on assets measured using market prices). However, it is not obvious that we can also capture the key time series pattern, which is that the wedge increased sharply from 2002 to 2016 while market concentration grew only modestly over the same time period. Using a simple calibration exercise, we show our model can indeed capture most of the observed stability in discount rates alongside the decline in risk-free rates and costs of capital. The reason is that the distortions induced by market power can be amplified by then secular increase in cross-sectional productivity dispersion even when market concentration is held fixed. When we also allow for a small increase in market concentration, we fully account for the lower weighted-average costs of capital in 2016, as well as a decline in real investment but an increase in cash holdings.

Setup. To illustrate the evolution of discount rates and investment wedges over time, we consider a simple overlapping generations version of the model. Time is discrete and indexed by $t = \{0, 1, 2, ...\}$. At each date, a new generation of young strategic agents and a competitive fringe enters the economy. Each strategic agent has additive time-separable

CRRA utility over consumption at dates t and $t+1$, and receives an initial endowment e. They can invest, consume, and trade Arrow securities in complete financial markets alongside a competitive fringe, and we consider the limit as in Section 3.4 in which the fringe is arbitrarily small ($m_f \to 0$) but the relative size $\frac{\mu}{m_f}$ converges to a constant, κ. Young agents make their investment and consumption decisions at date t and produce at date $t+1$. Old agents consume their output and exit the economy. With this setup, the decision problem of a strategic agent is identical to the one in equation (3).

There are four types of strategic agents (i.e., $N = 4$) indexed by their production technologies. There are green and red agents and, within each color type, volatile and stable firms. Consequently, we index an agent by $i \in \{gs, gv, rs, rv\}$ for "green stable", "green volatile", "red stable", and "red volatile", respectively. The stochastic process for risky investment returns for type i, $y_i(z)$, is composed of an aggregate component, Y_{z_A} for aggregate state $z_A \in \{L, H\}$, and an agent-specific component. The aggregate state follows a persistent two-state Markov process with transition matrix

$$\Pi = \begin{bmatrix} \rho_H & 1-\rho_H \\ 1-\rho_L & \rho_L \end{bmatrix}$$

where $\rho_H > \rho_L$. Agent-specific productivity is driven by a four-state process that is independent of aggregate productivity and across time, and whose realizations are boom, favorable, unfavorable, and busts. Favorable and unfavorable states each occur with probability 0.45, while boom and bust states each occur with probability 0.05. If a green agent receives a boom shock, there is a red agent who receives a bust shock, and vice versa. The same is true for favorable versus unfavorable shocks. Hence, these shocks are, in principle, perfectly diversifiable.

The difference between safe and volatile agents is safe agents have no exposure to booms and busts, while volatile firms have higher average productivity. We parameterize distributional production risk by $\delta(z_A)$, volatile production risk by $\Delta(z_A) > 0$, and average productivity gap by $b > 0$. The resulting state-contingent distribution of productivity for green agents is shown in Table 1. Since the productivity distribution of red agents is the mirror image, red and green firms of each type are ex-ante identical.

Due to the productivity gap b, it is immediate that only volatile firms invest in the competitive benchmark. This is because risk sharing among red and green agents is

	Green Agent Productivity			
	Boom	Favorable	Unfavorable	Bust
Volatile firm	$Y(z_A) + b + \delta + \Delta$,	$Y(z_A) + b + \delta$	$Y(z_A) + b - \delta$	$Y(z_A) + b - \delta - \Delta$
Stable firm	$Y(z_A) - b + \delta$,	$Y(z_A) - b + \delta$	$Y(z_A) - b - \delta$	$Y(z_A) - b - \delta$

Table 1: State-contingent productivity for green agents. Red agent productivity is the mirror image.

sufficient to eliminate all distributional risk. As a result, any investment by stable firms in our model is indicative of misallocation induced by market power.

Mapping model to data. We would like to speak to data on corporate discount rates and costs of capital. In the model, strategic agent i's discount rate, which is an internal hurdle rate, corresponds to the return on the wealth portfolio evaluated at her state prices (i.e., private valuations):

$$DiscountRate_i = \frac{\sum_z \pi(z)\, c_{2,i}(z)}{e - c_{1,i} - \sum_z \Lambda_i(z)\,(y_i(z)\, k_i + Rs_i) - k_i - s_i}. \tag{14}$$

The perceived WACC is the analogous expression, but evaluated at market prices:

$$WACC_i = \frac{\sum_z \pi(z)\, c_{2,i}(z)}{e - c_{1,i} - \sum_z q(z)\,(y_i(z)\, k_i + Rs_i) - k_i - s_i}, \tag{15}$$

We construct cross-sectional averages of these moments by weighting with model-implied market values of each firm type.

Experiment. To explore trends in rates of return and investment, we conduct the following thought experiment. First, we fit our model to the risk-free rate and corporate returns from Gormsen and Huber (2022) in 2002. One of the key parameters of the model is Δ, which determines the dispersion in firm-level productivity. We fix this parameter to match 1/2 times the interquartile range of 0.3258 estimated using Census data by Cunningham, Foster, Grim, Haltiwanger, Pabilonia, Stewart, and Wolf (2022), or 0.1629. We then feed in the 2016 level of Δ implied by the same paper, and ask to what extent our model matches the change in the risk-free rate and the rates of return reported by Gormsen and Huber (2022). We focus on the years 2002 to 2016 because this is the period of maximum overlap between the two data series. Appendix B provides details.

The model period is one year, and we interpret all returns to be nominal. We

choose the return process to be broadly in line with U.S. aggregate data. As discussed above, we set the distributional production risk Δ to be 0.1629. We choose the additional production volatility of volatile firms δ to be 0.10, which is approximately one quarter of the cross-sectional sales-growth volatility among Compustat firms. We set the productivity gap b to 0.01 and the gross return on storage to 1 (i.e., cash). The initial wealth of each strategic agent, e, is 3. To broadly measure the duration of booms and recession, we set the persistence of the high aggregate state, ρ_H to 2/3, and of the low aggregate state, ρ_L, to 0.10. We also set the aggregate productivity in the low state, Y_L, to 0.97.

Parameter	Interpretation	Value
R	Gross return to storage	1.000
e	Endowment of Strategic Agents	3.000
b	Productivity Gap	0.010
δ	Volatile Production Risk	0.100
Δ	Distributional Production Risk	0.163
ρ_H	Persistence of High State	0.666
ρ_L	Persistence of Low State	0.100
Y_L	Aggregate Productivity Low State	0.970
Y_H	Aggregate Productivity High State	1.243
γ, γ_f	Agent Risk Aversion	5.014
κ	Relative Size of Strategic Agents	0.393

Table 2: Parameter choices for the baseline calibration to 2002.

We calibrate the three remaining parameters using a simulated method of moments approach. Given our time frame, we focus on model-implied moments during booms only. We choose the aggregate productivity in the high state, Y_H, to match the 2002 1-year risk-free rate of 1.67%. We set the risk aversions of strategic agents and the competitive fringe to be the same (i.e., $\gamma = \gamma_f$), and target γ to match the 2002 perceived weighted average cost of capital (WACC) from Gormsen and Huber (2022) of 9.65%. Finally, we set the relative size of strategic agents κ to match the 2002 corporate discount rate from Gormsen and Huber (2022) of 15.20%. Table 2 reports the model parameters. We are able to achieve an exact correspondence between model and data. However, this fact should not be understood as model validation because we have more parameters than moments.

Moment	2002 Data	Model
Discount Rate	15.2%	15.20%
WACC	9.65%	9.65%
Risk-free Rate	1.67%	1.67%
Aggregate Investment	-	5.752
Aggregate Savings	-	0

Table 3: Data and model moments for the baseline calibration.

Moment	2016 Data	Experiment 1	Experiment 2
Discount Rate	16.47%	15.20%	15.06%
WACC	8.3%	8.56%	8.29%
Risk-free Rate	0.6%	0.63%	0.50%
Aggregate Investment	-	5.753	5.697
Aggregate Savings	-	0	0.030

Table 4: Data and model moments for the experiments. Experiment 1 considers an increase in Δ from 0.3258 to 0.3578. Experiment 2 additionally raises market concentration κ from 0.39 to 0.41

Interpreting Changes from 2002 to 2016. We now evaluate the extent to which increase in productivity dispersion and/or the degree of market concentration can account for the joint decline in the risk-free rate and costs of capital observed during the 2000s, as well as the decline in investment relative to firms' cost of capital and the increase in cash holdings. In our first experiment, we fix parameters at the 2002 values (given in Table 2), and exogenously feed in the change in productivity dispersion from Cunningham, Foster, Grim, Haltiwanger, Pabilonia, Stewart, and Wolf (2022), who estimate that the interquartile range in firm productivity rose from 0.3258 in 2002 to 0.3578 in 2016. This corresponds to a 10% increase in productivity dispersion. Importantly, this change is fully determined outside of our model.

Column 3 of Table 4 shows data and model-implied moments for 2016 given the exogenous change in dispersion. In line with the data, the risk-free rate falls to 0.63%, the WACC to declines to 8.57%, and the discount rate is unchanged at 15.20%. Investment does increase modestly for both safe and volatile firms, although it is small relative to the decline in the cost of capital. This is because the stark decline in the WACC weakly raises the value of individual production for both types despite worse risk sharing. Although these effects are not driven by *rising* market power, they are tightly related to the *level* of market concentration: if markets were perfectly competitive, equilibrium outcomes would be unaffected by dispersion.

To isolate the direct effects of market concentration, we conduct a second experiment, in which we modestly increase market concentration $\kappa = \frac{\mu}{m_f}$ from 0.39 to 0.41. Such a mild increase is broadly in line with the data from Kwon, Ma, and Zimmermann (2023), who show that market concentration was high but only modestly increasing starting from 2002. Column 4 of Table 4 shows our findings. A small increase in market concentration allows us to match the level of the WACC, and mildly lowers the risk-free rate and the discount rate. In contrast to the first experiment, risky investment declines and cash holdings (i.e., storage) increase relative to 2002 This is because the substitution effect induced by poor risk sharing now dominates the wealth effect.

Interestingly, our findings suggest it may be difficult to assess the full extent of distortions from market concentration using standard measures of price impact, such as price elasticities in financial markets. In particular, price elasticity in our model[13] is 0.324 for our baseline parameters, 0.351 under Experiment 1, and 0.362 under Experiment 2. Despite an increase in the underlying risk sharing friction, the elasticity is stable because both the quantity and price of risk is changing alongside trading volumes. This is consistent with Koijen and Yogo (2019), who estimate that price impact in equity markets is high but relatively stable over the time period in question.

Figures 1 and 2 illustrate the underlying mechanisms using comparative statics with respect to relative market concentration $\kappa = \frac{\mu}{m_f}$ and productivity dispersion Δ. Parameters are fixed at their 2016 values. The dashed vertical lines indicate the parameter values for 2016. The left panel of Figure 1 shows that, as concentration increases, all three rates of return fall. This in part because imperfect risk sharing raises the value of insurance, and in part because investment by the less productive type (the stable type) reduces the average return to investment (see also the right panel). The right panel shows that, when market concentration is sufficiently high, firms begin self-insuring through storage. This leads to a steeper decline in risky investment, flattens the decline in the risk-free rate, and accelerates the decline in risky rates of returns. When all types employ storage, the risk free rate must be equal the zero net return offered by storage.

Figure 2 considers changes in productivity dispersion Δ. All else equal, higher dispersion leads to higher trading needs. In concentrated markets, this raises the distortions from market power. This effect is mild as long as dispersion is not too high, in which

[13] For agent i and asset z, the elasticity is $\frac{q'(z)}{q(z)}\mu a_i(z)$. We report the cross-sectional average across i and z.

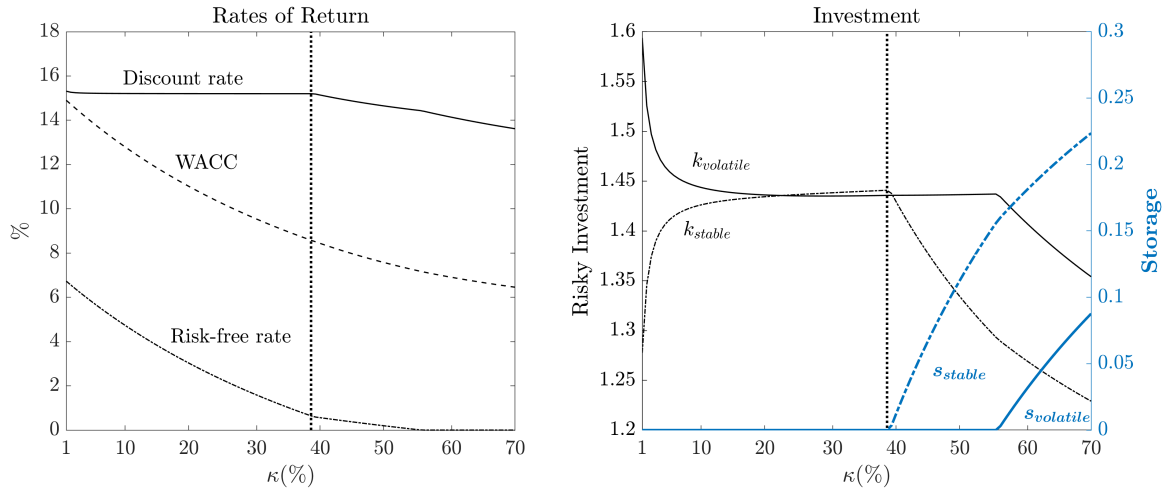

Figure 1: Comparative statics with respect to market concentration μ, where $\kappa = \mu/m_f$. The dispersion parameter Δ is set to its 2016 value. The dashed vertical line shows the calibrated value of κ for 2016. Values on the x-axis are rounded.

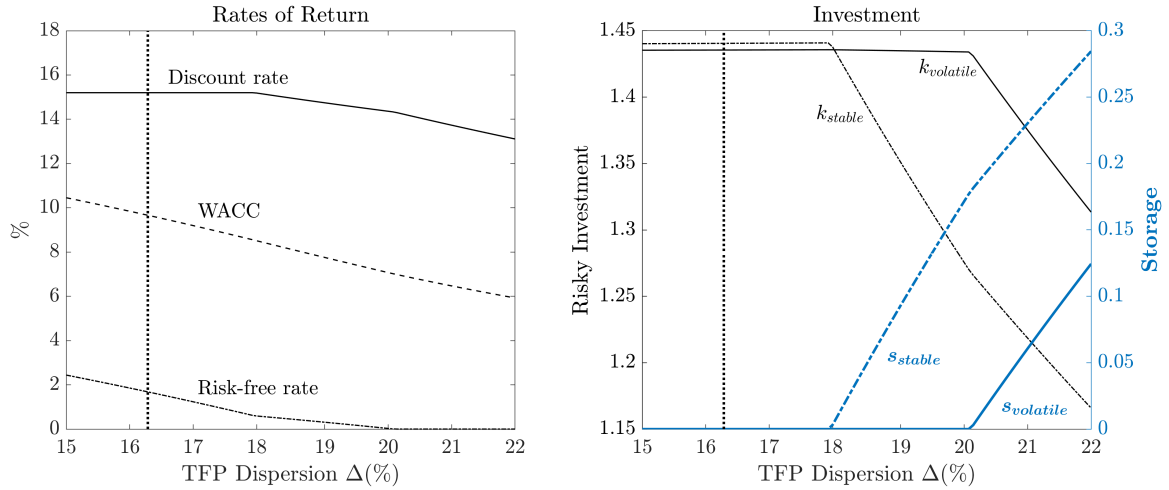

Figure 2: Comparative statics with respect to dispersion Δ. 2016 value of Δ shown in dashed vertical line. Values on the x-axis are rounded.

case agents continue to invest primarily in the risky technology. Since the distribution of income across states is thus approximately constant, the response is initially reflected through a rise in asset prices. This lowers the WACC and the risk-free rate. Once agents begin to employ storage, however, increasing productivity dispersion leads to a decline in the investment of both stable and volatile types, which reduces both output and production risk in the economy. This change in the quantity of risk and in future expected income further depresses all three rates of return.

Taken together, the model gives rise to a sharp nonlinearity in the relative response

of real quantities and returns. An increase in market concentration primarily exacerbates capital misallocation among strategic agents while an increase in productivity dispersion is principally reflected in inflated asset prices until storage is employed. Both lead to declines in not only the risk-free rate, but also in strategic agents' perceived WACC and discount rates. Such risk compression is qualitatively in line with recent data (e.g., Bianchi, Lettau, and Ludvigson (2020)).

5 Additional Empirical Implications

In addition to the evidence on firm wedges in discount rates and aggregate co-movement between investment, risk-free rates, and the cost of capital, our model makes several additional predictions. First, a necessary condition for our mechanism to operate is that large firms hedge but under-diversify their risks. That large firms hedge is well-documented in the literature (e.g., Allayannis and Weston (2001), Purnanandam (2008) Batram, Brown, Gregory, and Fehle (2009), Panaretou (2014)). Regarding under-diversification, Guay and Kothari (2003) provide evidence that 234 large non-financial corporations only minimally hedge against interest rate, exchange rate, and commodity price fluctuations with derivatives. Amel-Zadeh, Kasperk, and Schmalz (2022) also document that up to one-fifth of the largest U.S. firms have a single large nonfinancial blockholder or insider. A unique prediction of our theory is a negative relation between hedging activities and Tobin's Q.

Our theory also has connections to the production-based asset pricing literature, which uses the absence of arbitrage to relate the returns on a firm's equity and debt to the return on its assets. In the canonical theory (e.g., Cochrane (1996)), the return on assets is determined by how the firm values its marginal return on investment. In our theory, this relationship is modified by a wedge between the market's and an insider's valuation of a firm's assets. Notice we can express the first-order conditions from equation (8) as:

$$\sum_z R_I(z) - r_f^m = \underbrace{-Cov\left(\frac{q(z)}{\sum_z q(z)}, R_I(z)\right)}_{\text{Market Risk Premium}} + \underbrace{-\sum_z \frac{w_i(z)}{\sum_z q(z)} R_I(z)}_{\text{Strategic Trading Premium}} . \tag{16}$$

where $R_I(z) = \frac{y_i(z)k_i + Rs_i}{k_i + s_i}$ is the state-contingent return on investment. The second piece is a non-marketed excess return the firm garners from strategic trading in financial markets.

From equation (7), this piece is a function of price impact and a firm's financial positions, $a_i(z)$. Consequently, this piece is not a conventional risk premium, and is related to the firm's financial activities. Since it is difficult to measure a firm's financial positions in practice, characteristics that predict a firm's positions, such as those examined in Koijen and Yogo (2019), can proxy as factors to help explain cross-sectional equity excess returns.

6 Conclusion

We construct a model of concentrated financial markets in which large risk-averse firms invest in risky projects and internalize their price impact when trading state-contingent claims in financial markets. This results in cross-sectional misallocation of capital and aggregate under-investment that exacerbates financial market illiquidity in an adverse feedback loop. In line with recent trends, increased market concentration can lead to a joint decline in the risk-free rate, risk premia, investment, and productivity. Our analysis can also explain the documented wedge between firm investment hurdle rates and financial market returns, and how it varies with the risk-free rate, as well as the rise of cash holdings among nonfinancial corporations. Our framework is tractable and useful for studying the economic consequences of financial market concentration more generally.

References

AI, H., K. LI, AND F. YANG (2020): "Financial Intermediation and Capital Reallocation," *Journal of Financial Economics*, 138, 663–686.

ALEXANDER, L., AND J. EBERLY (2018): "Investment Hollowing Out," *IMF Economic Review*, 6, 5–30.

ALLAYANNIS, G., AND J. P. WESTON (2001): "The Use of Foreign Currency Derivatives and Firm Market Value," *Review of Financial Studies*, 14, 246–276.

ALLAZ, B. (1992): "Oligopoly, Uncertainty and Strategic Forward Transactions," *International Journal of Industrial Organization*, 10, 297–308.

ALLAZ, B., AND J.-L. VILA (1993): "Cournot Competition, Forward Markets and Efficiency," *Journal of Economic Theory*, 59, 1–16.

AMEL-ZADEH, A., F. KASPERK, AND M. SCHMALZ (2022): "Mavericks, Universal, and Common Owners - The Largest Shareholders uf US Public Firms," *Working Paper*.

ASPLUND, M. (2002): "Risk-averse Firms in Oligopoly," *International Journal of Industrial Organization*, 20, 995–1012.

AZAR, J., AND X. VIVES (2021): "General Equilibrium Oligopoly and Ownership Structure," *Econometrica*, 89, 999–1048.

BAO, J., J. PAN, AND J. WANG (2011): "The Illiquidity of Corporate Bonds," *Journal of Finance*, 66, 911–46.

BATRAM, S., M. BROWN, W. GREGORY, AND F. R. FEHLE (2009): "International Evidence on Financial Derivatives Usage," *Financial Management*, 38, 185–206.

BEN-DAVID, I., F. FRANZONI, R. MOUSSAWI, AND J. SEDUNOV (2021): "The Granular Nature of Large Institutional Investors," *Management Science*, 67(11), 6629–6659.

BIANCHI, F., M. LETTAU, AND S. C. LUDVIGSON (2020): "Monetary Policy and Asset Valuation," *mimeo Duke, UC Berkeley, and NYU*.

BIGIO, S. (2015): "Endogenous Liquidity and the Business Cycle," *American Economic Review*, 105, 1883–1927.

BOLTON, P., H. CHEN, AND N. WANG (2011): "A Unified Theory of Tobin's q, Corporate Investment, Financing, and Risk Management," *Journal of Finance*, 66, 1545–1578.

BRETSCHER, L., L. SCHMID, I. SEN, AND V. SHARMA (2022): "Institutional Corporate Bond Pricing," *Working Paper*.

CARLSTROM, C. T., AND T. S. FUERST (1997): "Agency Costs, Net Worth, and Business Fluctuations: A computable General Equilibrium Analysis.," *American Economic Review*, 87, 893–910.

CARVAJAL, A. (2018): "Arbitrage pricing in non-Walrasian financial markets," *Economic Theory*.

CELIK, S., G. DEMIRTAS, AND M. ISAKSSON (2020): "Corporate Bond Market Trends, Emerging Risks and Monetary Policy," *OECD Capital Market Series*.

CHEN, H., W. W. DOU, H. GUO, AND Y. JI (2020): "Feedback and Contagion Through Distressed Competition," *Working Paper*.

CHEN, P., L. KARABARBOUNIS, AND B. NEIMAN (2017): "The Global Rise of Corporate Savings," *Journal of Monetary Economics*, 89, 1–19.

COCHRANE, J. H. (1991): "Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations," *Journal of Finance*, 46, 209–237.

——— (1996): "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104, 572–621.

CORBAE, D., AND P. D'ERASMO (2021): "Capital Buffers in a Quantitative Model of Banking Industry Dynamics," *Econometrica*, 89(6), 2975–3023.

CORBAE, D., AND R. LEVINE (2018): "Competition, Stability, and Efficiency in Financial Markets," *Working Paper*.

CORHAY, A., H. KUNG, AND L. SCHMID (2020): "Competition, Markups and Predictable Returns," *Review of Financial Studies*, 33, 5906–5939.

COX, C., AND A. KARAM (2022): "Two-Period Duopolies with Forward Markets," *Review of Industrial Organization*, 60.

CROUZET, N., AND J. EBERLY (2021): "Intangibles, Markups, and the Measurement of Productivity Growth," *Journal of Monetary Economics*, 124, S92–S109.

CUNNINGHAM, C., L. FOSTER, C. GRIM, J. HALTIWANGER, S. W. PABILONIA, J. STEWART, AND Z. WOLF (2022): "Dispersion in Dispersion: Measuring Establishment-Level Differences in Productivity," *BLS Working Paper*.

DE GIOVANNI, D., AND E. IAKIMOVA (2022): "Strategic Capacity Choice with Risk-averse Firms," *Journal of the Operational Research Society*, 0, 1–17.

DEL NEGRO, M., D. GIANNONE, M. P. GIANNONI, AND A. TAMBALOTTI (2018): "Global Trends in Interest Rates," *FRBNY Staff Report*, 866.

DRECHSLER, I., A. SAVOV, AND P. SCHNABL (2017): "The Deposits Channel of Monetary Policy," *The Quarterly Journal of Economics*, 132(4), 1819–1876.

DRÉZE, J. (1974): "Investment under Private Ownership: Optimality, Equilibrium and Stability," in *Allocation Under Uncertainty*. New York:Wiley.

EGAN, M., A. HORTACSU, AND G. MATVOS (2017): "Deposit Competition and Financial Fragility: Evidence from the US Banking Sector," *American Economic Review*, 107(1), 169–216.

EISFELDT, A. L., AND A. A. RAMIPINI (2006): "Capital Reallocation and Liquidity," *Journal of Monetary Economics*, 53, 369–399.

EISFELDT, A. L., AND Y. SHI (2018): "Capital Reallocation," *Annual Review of Financial Economics*, 10, 361–386.

FERNALD, J. G., R. E. HALL, J. H. STOCK, AND M. W. WATSON (2017): "The Disappointing Recovery of Output After 2009," *Technical report, Brookings Papers on Economic Activity*.

FERREIRA, M. H. (2021): "Aggregate Implications of Corporate Lending by Nonfinancial Firms," *Working Paper*.

GABAIX, X., AND R. KOIJEN (2020): "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," *mimeo Harvard and Chicago*.

GABSZEWICZ, J. J., AND J.-P. VIAL (1972): "Oligopoly "a la Cournot" in a General Equilibrium Analysis," *Journal of Economic Theory*.

GORMSEN, N., AND K. HUBER (2022): "Discount Rates: Measurement and Implications for Investment," *Working Paper, University of Chicago*.

GREENWALD, B., AND J. STIGLITZ (1990): "Asymmetric Information and the New Theory of the Firm: Financial Constraints and Risk Behavior," *American Economic Review*, 80, 160–165.

GRUBER, J. W., AND S. B. KAMIN (2015): "The Corporate Saving Glut in the Aftermath of the Global Financial Crisis," *International Finance Discussion Papers*, 1150.

GUAY, W., AND S. KOTHARI (2003): "How Much do Firms Hedge with Derivatives?," *Journal of Financial Economics*, 70, 423–461.

GUTIERREZ, G., AND T. PHILIPPON (2017): "Investmentless Growth: An Empirical Investigation," *Brookings Papers on Economic Activity*, pp. 89–169.

GUTIÉRREZ, G., AND T. PHILIPPON (2017): "Investmentless Growth: An Empirical Investigation," *Brookings Papers on Economic Activity*, pp. 89–169.

HACHEM, K., AND Z. M. SONG (2021): "Liquidity Rules and Credit Booms," *Journal of Political Economy*, 129(10), 2721–2765.

HADDAD, V., P. HUEBNER, AND E. LOUALICHE (2021): "How Competitive is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing," *Working Paper*.

JAGANNATHAN, R., D. A. MATSA, I. MEIER, AND V. TARHAN (2015): "Why Do Firms use High Discount Rates?," *Journal of Financial Economics*, 120, 445–463.

JONES, C., AND T. PHILIPPON (2016): "The Secular Stagnation of Investment?," *Working Paper*.

KARABARBOUNIS, L., AND B. NEIMAN (2018): "Accounting for Factorless Income," *NBER Macro Annual 2018*.

KARGAR, M., B. LESTER, D. LINDSAY, S. LIU, P.-O. WEILL, AND D. ZÚÑIGA (2021): "Corporate Bond Liquidity during the COVID-19 Crisis," *Review of Financial Studies*, 34, 5352–5401.

KOGAN, L., AND D. PAPANIKOLAOU (2012): "Economic Activity of Firms and Asset Prices," *Journal of Monetary Economics*, 4, 361–384.

KOIJEN, R. S., AND M. YOGO (2019): "A Demand System Approach to Asset Pricing," *Journal of Political Economy*, 127, 1475–1515.

KURLAT, P. (2013): "Lemons Markets and the Transmission of Aggregate Shocks," *American Economic Review*, 103, 1463–1489.

KWON, S. Y., Y. MA, AND K. ZIMMERMANN (2023): "100 Years of Rising Corporate Concentration," *Working Paper*.

KYLE, A. S. (1989): "Informed Speculation with Imperfect Competition," *The Review of Economic Studies*, 56(3), 317–355.

LAUBACH, T., AND J. WILLIAMS (2016): "Measuring the Natural Rate of Interest Redux," *Business Economics*, 51(2), 57–67.

LI, J., AND H. YU (2022): "Investor Concentration, Liquidity and Bond Price Dynamics," *Working Paper*.

MA, Y. (2019): "Nonfinancial Firms as Cross-Market Arbitrageurs," *Journal of Finance*, 74(6), 3041–3087.

MALAMUD, S., AND M. ROSTEK (2017): "Decentralized Exchange," *American Economic Review*, 107(11), 3320–3362.

NEUHANN, D., AND M. SOCKIN (2023): "Wealth Dynamics and Financial Market Power," *mimeo UT Austin*.

NÆS, R., J. A. SKJELTORP, AND B. A. ØDEGAARD (2011): "Stock Market Liquidity and the Business Cycle," *Journal of Finance*, 66, 139–176.

PANARETOU, A. (2014): "Corporate Risk Management and Firm Value: Evidence from the UK Market," *European Journal of Finance*, 20, 1161–1186.

PAPANIKOLAOU, D., AND V. PANOUSI (2012): "Investment, Idiosyncratic Risk and Ownership," *Journal of Finance*, 67.

PURNANANDAM, A. (2008): "Financial Distress and Corporate Risk Management: Theory and Evidence," *Journal of Financial Economics*, 48, 706–739.

RADNER, R. (1974): "Note on Unanimity of Stockholders' Preferences among Alternative Production Plans: A Reformulation of the Ekern-Wilson Model," *Bell Journal of Economicsand Management Science*.

ROGNLIE, M. (2018): "Comment on "Accounting for Factorless Income" (Karabarbounis and Neiman, NBER Macro Annual 2018," *Working Paper, Northwestern University*.

ROSTEK, M., AND J. H. YOON (2020): "Exchange Design and Efficiency," *mimeo University of Wisconsin-Madison and University College London*.

SCHMID, L., J. DAVID, AND D. ZEKE (2020): "Risk-Adjusted Capital Allocation and Misallocation," *Working Paper*.

A Proofs of Propositions

A.1 Proof of Proposition 1:

Step 1: The Problem of the Fringe:

From the first-order condition for $a_f(z)$ from the competitive fringe's problem (2), we can recover the pricing equation of the Arrow-Debreu claim to security z

$$\tilde{q}(z) = \pi(z)u_f'\left(c_{2,f}(z)\right) = \Lambda_f(z),$$

where $\Lambda_f(z)$ is the competitive fringe's state price. Since $c_{2,f}(z) = e_{2,f}(z) + a_f(z)$, imposing the market-clearing condition, (1), reveals that

$$\tilde{q}(z) = \pi(z)u_f'\left(e_{2,f}(z) - \frac{1}{m_f}A(z)\right).$$

In equilibrium, this must be the realized price of the claim, $Q(\mathbf{A}, z)$. Consequently, the competitive fringe's Euler Equation pins down asset prices in the economy. As this price is a function of state variables from the perspective of the fringe, we designate the realized price more concisely as:

$$q(z) = Q(\mathbf{A}, z).$$

Step 2: Equilibrium Price Impact:

We next impose a consequence of our Cournot-Walras equilibrium concept. Since agents of type i take the demands of other agents (even within their type) as given. As a consequence,since $u_f(z)$ is twice continuously differentiable and each agent's position size scales by its mass μ, we can derive each agent's perceived price impact:

$$\frac{\partial \tilde{Q}_{j,i}(\mathbf{A}, z)}{\partial a_{j,i}(z)} = -\frac{\mu}{m_f}\pi(z)u_f''\left(c_{2,f}(z)\right) = -\frac{\mu}{m_f}\frac{\partial q(z)}{\partial A(z)},$$

which also implies that price impact is symmetric across all strategic agents. Defining $q'(z) = \frac{\partial q(z)}{\partial A(z)}$ yields the expression in the statement of the proposition.

Step 3: The Law of One Price:

The Law of One Price holds because the competitive fringe prices all assets. To see this, suppose there are two assets j and k with payoffs $x_j(z)$ and $x_k(z)$. Then:

$$q_j = \sum_{z \in \mathcal{Z}} x_j(z) \Lambda_f(z) = \sum_{z \in \mathcal{Z}} x_k(z) \Lambda_f(z) = q_j.$$

Since the fringe participates in all asset markets, no arbitrage is satisfied in our setting.

Step 4: Market Structure Invariance:

Suppose we have some arbitrary asset span indexed by the $|\mathcal{Z}| \times |\mathcal{Z}|$ matrix X that is of full rank. In the special case of Arrow-Debreu assets, $X = I_{|\mathcal{Z}|}$, i.e., the identity matrix of rank $|\mathcal{Z}|$. Let x_k index the kth row vector of X, and $x_k(z)$ be the dividend asset k pays in state z.

If the competitive fringe trades assets with asset span X, from the first-order conditions of the competitive fringe's optimization problem that the vector of asset prices $\vec{q}_X$ satisfies:

$$\vec{q}_X = X\vec{\Lambda}_f = X\vec{q}, \tag{17}$$

where $\vec{\Lambda}_f$ is the vector of the fringe's state prices and $\vec{q}$ the vector of Arrow asset prices.

The quasi-linear competitive fringe now maximizes $u_f\left(y_f(z) - \sum_{k=1}^{|\mathcal{Z}|} x(z) x_k(z) A_{x_k}(z)\right)$ $+ \sum_{k=1}^{|\mathcal{Z}|} x(z) q_{x_k} A_{x_k}(z)$, where $A_{x_k}(z)$ is the total demand for asset k of the strategic agents. It follows that the price impact function can be summarized by the matrix Γ:

$$\Gamma = XUX', \tag{18}$$

where U is the diagonal matrix with diagonal entries $-\frac{\mu}{m_f}\pi(z)u_f''\left(c_{2,f}(z)\right)$.

We now establish that whether the complete markets span is $I_{|\mathcal{Z}|}$ or X has no impact on allocations. Our arguments are similar in spirit to those in (Carvajal (2018)), but applied to our setting with production and do not impose quasi-linearity of strategic agents. If there are no real effects, the consumption allocations of the fringe, c_{f1} and $c_{2,f}(z)$, and its state prices, $\Lambda_f(z)$, must be the same in both economies.

Notice we can stack the first-order conditions for strategic agent i with asset span $I_{|\mathcal{Z}|}$ from equation (26) as:

$$\vec{\Lambda}_i = \vec{\Lambda}_f + U\vec{a}_i, \tag{19}$$

where $\vec{\Lambda}_i$ are the stacked state prices of agent i, $\vec{a}_i$ is the vector of her asset positions, and we have substituted for Arrow-Debreu prices $\vec{q}$ with $\vec{\Lambda}_f$.

Let $\vec{a_{i,x}}$ be the vector of asset positions of agent i when she instead trades with the asset span X. Imposing invariance of the consumption allocations of strategic agent i requires that:

$$\vec{a}_i = X'\vec{a_{i,x}}. \tag{20}$$

Substituting with equation (20), we can manipulate equation (19) to arrive at:

$$X\vec{\Lambda}_i = X\vec{\Lambda}_f + XUX'\vec{a_{i,x}} = X\vec{\Lambda}_f + \Gamma\vec{a_{i,x}}, \tag{21}$$

where we have also substituted with equation (18). This is the identical stacked first-order conditions if strategic agent instead traded asset span X.

Consequently, if the competitive fringe's consumption allocations are unchanged between asset spans, then so are the optimal portfolios of each strategic agent. If all strategic agents have the same asset demands, then their aggregate demand for asset exposures in each state z are the same. By market clearing, then, the state-specific asset exposures of the competitive fringe are the same in both asset spans, and consequently so are their consumption allocations, confirming our conjecture.

What remains to show is that capital and savings choices and budgets sets of strategic agents are unchanged across asset spans. This, however, is trivial because no arbitrage makes invariant the cost of state-specific asset exposures. Consequently, financing the same portfolio of state-specific asset exposures costs the same with asset span $I_{|\mathcal{Z}|}$ as with asset span X. Given the same capital and savings choices, k_i and s_i for each i, the marginal utility of each agent with market structure X is the same state-by-state as with market structure $I_{|\mathcal{Z}|}$, confirming that k_i and s_i are also optimal with market structure X.

A.2 Proof of Corollary 1:

Suppose the competitive fringe has convex marginal utility in addition to strictly concave utility. Then a decrease in its consumption at date 2 in state z not only increases its marginal utility but the derivative of its marginal utility in state z. From Proposition 1, this raises both the price of the Arrow security referencing state z and price impact in that market.

If total resources $Y(z)$ in state z fall because of lower productivity ($y_i(z)$), strategic agents' demand for insurance against state z from the fringe must weakly increase. This is because they have less aggregate resources with which to insure each other. If the fringe supplies sells more securities against state z, then it consumes less in state z and the claim in the corollary follows. If strategic agents invest less efficiently in risky capital and consume more at date 1, then there are again less aggregate resources among strategic agents at date 1, and the claim again follows.

A.3 Proof of Lemma 1:

Step 1: The Problem of Strategic Agents:

We first consider the optimization problem of strategic agent j of type i, (3). In what follows, we attach the Lagrange multiplier φ_i to the budget constraint. The first-order necessary conditions for $c_{1,i}, k_i, s_i$, and $\{a_i(z)\}_{z\in\mathcal{Z}}$ are then given by:

$$c_{1,j,i} \;:\; u'\left(c_{1,j,i}\right) - \varphi_{j,i} = 0, \tag{22}$$

$$k_{j,i} \;:\; \sum_{z\in\mathcal{Z}} \pi(z)\, u'\left(c_{2,j,i}(z)\right) y_i(z) - \varphi_{j,i} \leq 0 \;\left(= \; if\; k_{j,i} > 0\right), \tag{23}$$

$$s_{j,i} \;:\; \sum_{z\in\mathcal{Z}} \pi(z)\, u'\left(c_{2,j,i}(z)\right) R - \varphi_{j,i} \leq 0 \;\left(= \; if\; s_{j,i} > 0\right), \tag{24}$$

$$a_{j,i}(z) \;:\; \pi(z)\, u'\left(c_{2,j,i}(z)\right) - \varphi_{j,i}\left(\tilde{Q}_{j,i}(\mathbf{A}, z) + \frac{\partial \tilde{Q}_{j,i}(\mathbf{A}, z)}{\partial a_{j,i}(z)} a_{i,j}(z)\right) = 0. \tag{25}$$

The above represents the first-order necessary conditions for agent j of type i's problem. From (22), it is immediate that $\varphi_{j,i} = u'\left(c_{1,j,i}\right) \geq 0$ because marginal utility is nonnegative.

Now that we have derived the first-order necessary conditions for agent j of type i's optimal asset demands, we can impose the consistency required of a Cournot-Walras

equilibrium with the competitive fringe. Since strategic agent i has rational expectations, her perceived price impact must coincide with her actual price impact from (5) in Proposition (1). Consequently, the first-order necessary condition (25) reduce to:

$$a_{j,i}(z) : \Lambda_{j,i}(z) = q(z) + \frac{\mu}{m_f} q'(z) a_{j,i}(z) \, \forall \, z \in \mathcal{Z}, \tag{26}$$

where $\Lambda_{j,i}(z)$ is the state price of strategic agent j of type i in state z, i.e., $\Lambda_{j,i}(z) = \pi(z) \frac{u'(c_{2,j,i}(z))}{u'(c_{1,j,i})}$.

We next establish that the correspondence for admissible controls from the constraint set of strategic agent j of type i is compact-valued.

Notice first that strategic agent j, i would never take an infinite position in any asset. If a strategic agent takes an infinite negative position in asset z, $a_{j,i}(z) \to -\infty$, then there are two cases to consider for the right-hand side of (26). First, if another strategic agent takes an off-setting position in the asset, then $q(z)$ and $q'(z)$ remain positive and well-defined, and the right-hand side tends to $-\infty$ because of the $a_{j,i}(z)$ term. Second, if the fringe is forced to absorb the supply, then $q(z)$ falls because the fringe's state prices are decreasing in the fringe's consumption. Either the prices remain positive, in which case, the previous conclusion that the left-hand side is $-\infty$ holds, or prices tend to zero and the left-hand side tends to zero. In both cases, the left-hand side remains positive and may tend to ∞ because initial consumption $c_{1,i}$ becomes infinite, $c_{1,i} \to \infty$. This is clearly a contradiction as the seller would not want to be a buyer in that security market.

A similar argument applies to infinite demand, in which case the right-hand side of (26) tends to positive ∞ (the demand is either offset by a strategic agent or absorbed by the fringe through infinitely negative date 2 consumption in state z). The left-hand side, however, tends to zero with infinite consumption at date 2, which contradicts the equality of the first-order condition. As such, no strategic agent will take an infinite position in any security.

Notice next that the capital and storage choices by strategic agent j, i, $k_{j,i}$ and $s_{j,i}$, respectively, are also bounded. First, they are restricted to be nonnegative by feasibility. Second, because no agent would ever take an arbitrarily negative asset position, the total resources available for capital and storage are consequently also bounded.

Finally, consumption at both dates is bounded. At date 2, this is the case because endowments and production payoffs are bounded, and storage and capital decisions are

also bounded. At date 1, this is the case because all security positions are bounded.

Consequently, we can bound all controls of strategic agent j, i's problem, $\{c_{1,j,i}, \{a_{j,i}(z)\}_{z \in \mathcal{Z}}, k_{j,i}, s_{j,i}\}$, in a closed and bounded set. By the Heine-Borel Theorem, this set is compact.

We now recall from Proposition (1) that the pricing functional $Q_{j,i}(\mathbf{A}, z)$ is continuously differentiable in $\mathbf{A}$ because it is the marginal utility of the competitive fringe in state z, $\pi(z)\, u'_f\left(c_{2,f}(z)\right)$. Since the state prices of the strategic agents and the price impact functional are continuous because all utility functions are $\mathcal{C}^2$, strategic agent j, i's choice correspondence set is also continuous in the optimization problem's primitives (i.e., production processes and initial endowments). As such, the choice correspondence of strategic agent j, i's problem is continuous and compact-valued.

It then follows because the objective function of strategic agent j, i is continuous (in fact, differentiable), and the choice correspondence is continuous and compact-valued, that by Berges' Theory of the Maximum a solution to the decision problem of strategic agent j, i exists. As the choice of j, i was arbitrary, this holds for all agents j of type i and all types $i \in \{1, ..., N\}$.

Step 2: Existence:

As a result of Berge's Theory of the Maximum, the optimal policies of each strategic agent are upper-hemicontinuous correspondences. We can then construct a mapping from a conjectured set of investment and asset decisions for all strategic agents to an optimal set of investment and asset decisions using the market-clearing conditions (1) and the optimal policy correspondences as an equilibrium correspondence whose image is a compact space. Since the budget constraints of strategic agents are not necessarily convex because of price impact, we allow for randomization of consumption bundles to ensure that the compact space is also convex. We can then apply Kakutani's Fixed Point Theorem to conclude that an equilibrium exists.

Step 3: Homogeneity of Optimal Policies in Initial Wealth:

Suppose that the optimal policies of strategic agent j of type i satisfy $c_{1,j,i} = \hat{c}_{j,i,1} e$,

$c_{2,j,i}(z) = \hat{c}_{2,j,i}(z)\, e$, $k_{j,i} = \hat{k}_{j,i} e$, $s_{j,i} = \hat{s}_{j,i} e$, and $a_{j,i}(z) = \hat{a}_{j,i}(z)\, e$. We then rewrite the FONCs (23), (24), and (26) for strategic agent j of type i, given the homotheticity of strategic agent preferences as:

$$\hat{k}_{j,i} \;:\; \sum_{z \in \mathcal{Z}} \pi(z) \frac{u'\left(\hat{c}_{2,j,i}(z)\right)}{u'\left(\hat{c}_{1,j,i}\right)} y_i(z) - 1 \leq 0 \;\left(= \; if\; k_{j,i} > 0\right), \tag{27}$$

$$\hat{s}_{j,i} \;:\; \sum_{z \in \mathcal{Z}} \pi(z) \frac{u'\left(\hat{c}_{2,j,i}(z)\right)}{u'\left(\hat{c}_{1,j,i}\right)} R - 1 \leq 0 \;\left(= \; if\; s_{j,i} > 0\right), \tag{28}$$

$$\hat{a}_{j,i}(z) \;:\; \pi(z) \frac{u'\left(\hat{c}_{2,j,i}(z)\right)}{u'\left(\hat{c}_{1,j,i}\right)} - q(z) - \frac{\mu}{m_f} \hat{q}'(z) \hat{a}_{j,i}(z) = 0, \tag{29}$$

where we recognize that $\hat{q}'(z) = \frac{1}{e} q'(z)$, where $\hat{q}'(z) = \frac{\partial \tilde{Q}_{j,i}(\mathbf{A}, z)}{\partial \hat{a}_{j,i}(z)}$. It then follows that, conditional on prices $q(z)$, the optimal policies of strategic agent j of type i are indeed homogeneous of degree 1 in e.

A.4 Proof of Corollary 2:

It is immediate if the competitive fringe has a lower endowment $e_f(z)$ in state z, then it also consumes (weakly) less in state z. This raises the price of the Arrow security referencing state z $q(z)$ and with convex marginal utility, price impact $\frac{\mu}{m_f} q'(z)$. This is because the Arrow price is equal to the fringe's marginal utility in that state from Proposition 1.

Holding fixed the investment policies of strategic agents for the moment, the rise in the asset price $q(z)$ implies a larger gap between $\Lambda_i(z)$ and $q(z)$ for seller i, i.e., $\Lambda_i(z) - q(z) = \frac{\mu}{m_f} q'(z)\, a_i(z)$ from the first-order condition for $a_i(z)$ in Proposition 1 becomes more negative. This may, however, involve more selling of securities $a_i(z)$ despite the rise in price impact $\frac{\mu}{m_f} q'(z)$. This raises i's investment wedge in state z $w_i(z) = \frac{\mu}{m_f} q'(z)\, a_i(z)$ in state z. From Proposition 1, this increase distorts his investment choice further from its competitive choice of investment.

As such, there is more capital misallocation by sellers in Arrow market z.

A.5 Proof of Corollary 3:

First fix the capital investment of all agents, $k_i, s_i \; \forall \; i$, and consider the competitive equilibrium in which $\mu = 0$. Suppose we alter agents' production technologies to reduce

productivity in states in which agents overlap in production and increase it in states in which they do not, such that total output in each state remains unchanged. Let the new productivities be indexed by $\{\tilde{y}_i(z)\}_{i=1}^N$, and the new asset positions from retrading after the redistribution be $a_i(z)$. For instance, with two agents and two states of production, we can shift productivity so that agent i now produces all output in state 1, $y_i(1) + \frac{k_{i'}}{k_i} y_{i'}(1)$, and agent j produces all output in states 2, $y_{i'}(2) + \frac{k_i}{k_{i'}} y_i(2)$.

Because agents can insure each other against states in which they differ in production compared to states in which they jointly produce, gains from trade increase in the economy. With perfect competition, agents would trade until state prices are equalized, and consequently the redistribution of productivity is irrelevant to the equilibrium consumption allocation. The trading volume in their asset positions for claims in state z is then $\sum_{i=1}^N |\tilde{a}_i(z) - a_i(z)|$, where $\tilde{a}_i(z) = c_{1,i}(z) - \tilde{y}_i(z) k_i - Rs_i$ and $c_{2,i}(z)$ is agent $i's$ consumption in the competitive equilibrium.

Suppose instead agents are strategic. Because agents now internalize their price impact, however, they ration their asset demands and supplies to tilt prices in their favor, $\frac{\partial q(z)}{\partial A(z)} \Delta a_i(z)$ for a change in position of $\Delta a_i(z) = \tilde{a}_i(z) - a_i(z)$, from manipulating prices. As such, total trading volume is bounded from above by $\sum_{z \in \mathcal{Z}} \sum_{i=1}^N |\Delta a_i(z)|$, which are positive related to total gains from trade. Because agents can always choose to trade fully their differences in risk exposures, the distortions must be (weakly) larger with this increase in pure agent-specific risk.

A.6 Proof of Lemma 2:

The claim is immediate from inspection of the first-order conditions for strategic agent i's optimal portfolios from Lemma 1. In markets where $\Lambda_i(z) > q(z)$ (i.e., a buyer), the wedge is positive $\frac{\mu}{m_f} \hat{q}'(z) \hat{a}_{j,i}(z) > 0$. Similarly, in markets where i is a seller, or $\Lambda_i(z) < q(z)$, then the wedge is negative $\frac{\mu}{m_f} \hat{q}'(z) \hat{a}_{j,i}(z) < 0$. With declining marginal utility, an agent has low state prices relative to asset prices in states in which he has high output, and high state prices relative to asset prices in states in which he has low output.

With perfect competition, all agents equate their state prices with asset prices state-by-state. As such, the aforementioned wedges reflect the consequences of price impact.

A.7 Proof of Proposition 2:

Let k_i^{CE} be the scale of capital an agent would choose in the competitive equilibrium without price impact. Further, let k_i^{Aut} be the scale of capital an agent would choose in autarky. Two forces impact the choice of capital with price impact. The first is that all agents are more exposed to their own production than in the competitive environment because they trade less. This reduced risk sharing distorts the investment choices of all agents toward their autarky values. The second is that sellers sell less assets, which lowers their state prices relative to the competitive equilibrium, while buyers buy less assets, which raises their state prices relative to it. This depresses the investment of agents that are systematic sellers across security markets and raises the investment of systematic buyers.

Suppose now agent i chooses its capital such that $k_i^{CE} > k_i^{Aut}$. As a result of impaired risk sharing, it chooses a lower scale of production in the market equilibrium, converging to its autarky value when markets are sufficiently concentrated. If instead $k_i^{CE} \leq k_i^{Aut}$, then we have the opposite result. The agent chooses a higher scale of production in the market than in the competitive equilibrium, converging to its autarky value when markets are sufficiently concentrated.

For the third part, because total risky investment eventually converges to its autarky value, in which it is lower, for μ sufficiently large, total risky investment declines relative to the competitive equilibrium because of market concentration. Since price impact acts as an effective tax on the joint production among agents, capital misallocation rises, which lowers average productivity.

For the final part, if inefficient storage is used in the competitive equilibrium in which agents perfectly share risk, then it is also used in the market equilibrium and autarky. The converse, however, is not true: if storage is employed under the market, it need not be employed in the competitive equilibrium because the latter is efficient. As such, storage is always (weakly) higher in the market than in the competitive equilibrium.

A.8 Proof of Lemma 3:

We define the market risk premium RP_{mkt} as $\frac{E[\sum_i y_i(z)k_i]}{\sum q(z)\sum_i y_i(z)k_i} - r_m$. Direct manipulation of the FOCs for optimal investment and asset holdings in state z from Proposition 1 reveals

$$\begin{aligned}\sum_{z\in\mathcal{Z}} q(z)\, y_i(z)\, k_i &= \sum_{z\in\mathcal{Z}} q(z)\, E\left[y_i(z)\, k_i\right] + Cov\left(q(z), y_i(z)\, k_i\right) \\ &= k_i - \frac{\mu}{m_f}\sum_{z\in\mathcal{Z}} q'(z)\, a_i(z)\, y_i(z)\, k_i,\end{aligned}$$

from which follows that

$$\begin{aligned}\frac{E\left[\sum_i y_i(z)\, k_i\right]}{\sum_{z\in\mathcal{Z}} q(z)\sum_i y_i(z)\, k_i} - \frac{1}{\sum_{z\in\mathcal{Z}} q(z)} &= -\frac{Cov\left(\frac{q(z)}{\sum q(z)}, \sum_i y_i(z)\, k_i\right)}{\sum_{z\in\mathcal{Z}} q(z)\sum_i y_i(z)\, k_i} \\ &\quad - \frac{Cov\left(\frac{q(z)}{\sum_{z\in\mathcal{Z}} q(z)}, \sum_i y_i(z)\, k_i\right)}{\sum_i k_i - \frac{\mu}{m_f}\sum_{z\in\mathcal{Z}} q'(z)\sum_i a_i(z)\, y_i(z)\, k_i}.\end{aligned}$$

Given that $\frac{1}{\sum_{z\in\mathcal{Z}} q(z)}$ is the inverse of the market-implied riskless rate r_m, we arrive at the statement in the proposition.

A.9 Proof of Proposition 3:

Step 1: Asset prices compared a pseudo-competitive equilibrium:

For now, let us fix all investment decisions from the market equilibrium to be the investment decisions in the competitive equilibrium. Denote this pseudo-competitive equilibrium by the superscript $CE1$. This is equivalent to assuming all agents are in an endowment economy. We will return to the impact of market concentration on production in the sequel. We first consider the case in which agents do not employ storage in the market equilibrium.

We start with the aggregated FOCs (11):

$$q(z) + \frac{1}{N}\frac{\mu}{m_f} q'(z)\, A(z) = \mathbb{E}^*\left[\Lambda_i(z)\right]. \tag{30}$$

Consider the net demand of strategic agents $A(z)$. Suppose that net demand is weakly greater than under the competitive equilibrium, $A^{CE1}(z)$, i.,e., $A(z) \geq A^{CE1}(z)$ for all

security markets. Then, because $q(z) = \pi(z)u'_f\left(e_{2,f}(z) - \frac{1}{\mu}A(z)\right)$ from Proposition (1), it follows that $q(z) \geq q^{CE1}(z)$, and we are done.

We therefore focus on the case in which $A(z) < A^{CE1}(z)$ and attempt to establish a contradiction. Notice, if $A(z) < A^{CE1}(z)$, then from (30):

$$q(z) - q^{CE1}(z) = \mathbb{E}^*[\Lambda_i(z)] - \Lambda^{CE1}(z) - \frac{1}{N}\frac{\mu}{m_f}q'(z)A(z), \tag{31}$$

where $E^*\left[\Lambda_i^{CE1}(z)\right] = \Lambda^{CE1}(z)$ because state prices are all aligned in the competitive equilibrium (Lemma 1).

We now make use of Assumption 1 that the competitive fringe's trading positions in the competitive equilibrium satisfy $a_f(z) \geq 0$, which implies by market-clearing (1) that $A^{CE}(z) \leq 0$. This consequently also implies that $A(z) < A^{CE}(z) \leq 0$ given our focus on the case in which $A(z) < A^{CE}(z)$. Imposing this observation in (31) implies:

$$q(z) - q^{CE1}(z) \geq \mathbb{E}^*[\Lambda_i(z)] - \Lambda^{CE1}(z). \tag{32}$$

Suppose all types are symmetric (i.e., productivity risk is such that all types face symmetric problems). Then all types make the same investment decisions, $k_i = k$ and $s_i = s$, and total asset expenditures $\sum_{z\in\mathcal{Z}} q(z)a(z)$. It is then apparent that initial consumption is the same $c_{1i} = c_1$.

In this case, we can make use of the assumption that strategic agent marginal utility $u'(\cdot)$ is homothetic and strictly convex to apply Jensen's Inequality:

$$\mathbb{E}^*[\Lambda_i(z)] = \pi(z)\,\mathbb{E}^*\left[u'\left(\frac{c_{2i}(z)}{c_{1i}}\right)\right] \geq \pi(z)\,u'\left(\mathbb{E}^*\left[\frac{c_{2i}(z)}{c_{1i}}\right]\right). \tag{33}$$

Since $c_{1i} = c_1$ is the same across all agents, (33) reduces to:

$$\mathbb{E}^*[\Lambda_i(z)] \geq \pi(z)\,u'\left(\frac{\frac{1}{N}\sum_{i=1}^N c_{2i}(z)}{c_{1i}}\right) = \pi(z)\,u'\left(\frac{\sum_{i=1}^N y_i(z)k + A(z)}{N(e-s-k) + m_f(e-c_{1,f})}\right). \tag{34}$$

because:

$$Nc_1 = N(e-s-k) + m_f(e-c_{1,f}),$$

by market clearing.

Since $A(z) < A^{CE1}(z)$, by assumption, we recognize that:

$$\frac{\sum_{i=1}^{N} y_i(z)\,k + A(z)}{N(e-s-k) + m_f\left(e - c_{1,f}\right)} < \frac{\sum_{i=1}^{N} y_i(z)\,k + A^{CE1}(z)}{N(e-s-k) + m_f\left(e - c_{1,f}^{CE1}\right)}, \tag{35}$$

because $A(z) < A^{CE1}(z)$ also implies that resources are transferred from the second to first period through asset purchases by the competitive fringe.

The following Lemma characterizes state prices in the competitive equilibrium.

Lemma 4 *State prices in the competitive equilibrium, $\Lambda^{CE}(z)$ satisfy*

$$\Lambda^{CE}(z) = \pi(z)\,u'\left(\frac{\sum_{i=1}^{N} y_i(z)\,k + A^{CE}(z)}{N(e-s-k) + m_f\left(e - c_{1,f}^{CE}(z)\right)}\right).$$

Given that marginal utility is decreasing in consumption growth with homothetic preferences, Lemma 1 and (35) implies that:

$$\mathbb{E}^*\left[\Lambda_i(z)\right] \geq \pi(z)\,u'\left(\frac{\sum_{i=1}^{N} y_i(z)\,k + A^{CE1}(z)}{N(e-s-k) + m_f\left(e - c_{1,f}^{CE1}\right)}\right) = \Lambda^{CE1}(z). \tag{36}$$

Consequently, substituting (36) into (30) reveals:

$$q(z) \geq q^{CE1}(z), \tag{37}$$

which implies $A(z) \geq A^{CE1}(z)$ from Proposition (1), which is a contradiction.

This establishes that asset prices are higher state-by-state compared to a pseudo-competitive economy in which we fixed all investment decisions to be the same as in the market equilibrium. This is true when agents do not employ storage.

Suppose now agents employ storage. Since they are symmetric across types, they either all will use storage or they all will not. For an agent who uses storage, the sum of his state prices is fixed at $\frac{1}{R}$, or

$$E\left[\Lambda_i(z)\right] = \frac{1}{R} \forall i \in 1, ..., N.$$

Because only the sum of his state prices are constrained, state prices still are higher by Jensen's Inequality. If all agents employ storage, however, then the sum of all state prices

is constrained, and so is the cross-sectional average of the sum. Then, because average state prices cannot all be higher than in the pseudo-competitive equilibrium, the cross-sectional average must cease to change; otherwise, the sum would exceed $\frac{1}{R}$, a contradiction. In this case, the state prices of high marginal utility states are higher while those of low marginal utility states are lower, so that their sum is fixed at $\frac{1}{R}$. Consequently, while all Arrow-Debreu prices initially rise with market concentration, those of low marginal utility states start to fall when all agents employ storage. Although they fall, they remain elevated above their competitive equilibrium.

Step 2: Asset prices compared to the competitive equilibrium:

Since we have established that state prices are (weakly) higher with market concentration than in an equivalent endowment economy with the same capital and savings decisions, what remains is to compare the competitive equilibrium to this pseudo-competitive equilibrium with perfect risk sharing but the capital allocation decisions from the market equilibrium. Then, our claim will follow by transitivity.

It is immediate that the competitive equilibrium achieves the first-best allocation by the First Welfare Theorem. The competitive equilibrium is consequently equivalent to solving the Planner's problem:

$$\begin{aligned} U_0 &= \sup_{\{k_i, s_i, c_{1i}, c_{2i}(z)\}_{i=1}^N, s, c_{f1}, c_{2,f}(z)} \mathbb{E}\left[\sum_{i=1}^N u\left(c_{1,i}\right) + u\left(c_{2,i}\left(z\right)\right) - c_{f1} + u_f\left(c_{2,f}\left(z\right)\right)\right] \\ s.t. &: \sum_{i=1}^N c_{1,i} + c_{f1} + k_i + s = Ne + m_f e, \\ &: \sum_{i=1}^N c_{2,i}\left(z\right) + m_f c_{2,f}\left(z\right) = m_f e_{2,f}\left(z\right) + \sum_{i=1}^N y_i\left(z\right) k_i + Rs, \end{aligned}$$

where s is the aggregate storage.

The solution to this problem is characterized in Lemma 1, in which there is a unique social state price $\Lambda^{CE}\left(z\right)$ that determines all consumption sharing and production decisions. By construction, the distribution of investment with market concentration can achieve no higher utility than under the optimal centralized policy with the same resource constraint. As such, there exist improvements that (weakly) raise $c_{2,i}\left(z\right)$ in all

states by shifting investment away from less toward more productive technologies

From Proposition 2, agents that would invest in capital in the first-best under-invest in the noncompetitive economy because of strategic frictions, while those that do not invest may start to invest because of diminished opportunities to finance the production of other agents. For the same resources transferred intertemporally, $\sum_{i=1}^{N} e - c_{1,i}$, the first-best employs more efficient technologies without cross-sectional misallocation. As such, while some agents may under-invest and others over-invest with price impact, aggregate investment and investment efficiency falls. As such, state prices are (weakly) higher in all states in the pseudo-competitive equilibrium.

Finally, we consider the role of storage. If storage is used in the competitive equilibrium, then it is also employed in the market equilibrium, although the converse need not be true. When storage is employed by all agents, then the sum of their state prices is constrained to be $\frac{1}{R}$. As such, since capital allocation is less efficient in the market equilibrium, and the sum of state prices is constrained, it follows that prices increase for high marginal utility and decrease for lower marginal utility states to leave the average unchanged. If there is storage in the competitive equilibrium, then this effect is there for all μ, otherwise it becomes operative once all agents employ storage. It then follows that:

$$\Lambda^{CE1}(z) \geq \Lambda^{CE}(z),$$

as required.

Step 3: Risk-free Rate:

The second part of the claim follows directly Steps 1 and 2. Because $q(z)$ is higher state-by-state than in the competitive equilibrium, it follows that $\sum_{z \in \mathcal{Z}} q(z)$ is also larger than in the competitive equilibrium. Because r_f^m is the inverse of the sum of the state prices, the claim then follows. Moreover, $r_f^m \geq R$ because storage can be traded without market impact.

The floor of r_f^* is also R because, from the FOCs in Proposition 1, if there is storage, then $E\left[u_2'\left(\frac{c_{2,i}(z)}{c_{1,i}}\right)\right] = \frac{1}{R}$. To see that this is a lower bound, a necessary condition that agent i to invest in capital in its production technology is $E[y_i(z)] > R$, with the strict inequality necessary to embed a risk premium for the agent. Consequently, as long as

agent i holds storage, then $E\left[u_2'\left(\frac{c_{2,i}(z)}{c_{1,i}}\right)\right] = \frac{1}{R}$, and only once it exhausts all its resources in state contingent claims and capital, then $E\left[u_2'\left(\frac{c_{2,i}(z)}{c_{1,i}}\right)\right] < \frac{1}{R}$. Because this holds for all agents, it follows that $r^* \geq R$.

Step 4: Market Risk Premium:

In the special case in which strategic agents are symmetric across types, all types choose the same level of capital $k_i = k$, and the market risk premium (12) reduces to:

$$\begin{aligned} RP_{mkt} &= \frac{E\left[\sum_i y_i(z)\right]}{\sum_{z\in\mathcal{Z}} q(z) \sum_i y_i(z)} - r_m \\ &= r_m \frac{\sum_{z\in\mathcal{Z}} q(z) E\left[\sum_i y_i(z)\right] - \sum_{z\in\mathcal{Z}} q(z) \sum_i y_i(z)}{\sum_{z\in\mathcal{Z}} q(z) \sum_i y_i(z)} \\ &= -r_m \frac{Cov\left[q(z), \sum_i y_i(z)\right]}{E\left[q(z) \sum_i y_i(z)\right]}. \end{aligned}$$

It is immediate that the market-implied risk-free rate r_m is lower from Step 3 and that asset prices $q(z)$ are higher in the denominator in the market compared to the competitive equilibrium (from Step 2). In addition, and more subtle, is that the covariance between state prices and aggregate productivity, $Cov\left[q(z), \sum_i y_i(z)\right] < 0$, is less negative because all asset prices are inflated to reflect market concentration beyond underlying risk. Consequently, the market risk premium is lower in this special case.

A.10 Proof of Lemma 4:

In this lemma, we characterize the competitive equilibrium without market concentration. The standard first-order conditions for optimal consumption and asset holdings align state prices for all agents state-by-state:

$$q(z) = \frac{\pi(z) u'(c_{2i}(z))}{u'(c_{1i})} = \pi(z) u_f'(c_{2,f}(z)) = \Lambda^{CE}(z), \tag{38}$$

which implies for the N types of agents with homothetic preferences:

$$\frac{c_{2i}(z)}{c_{1i}} = \frac{c_{2j}(z)}{c_{1j}} = \eta(z), \tag{39}$$

and for the competitive fringe:

$$c_{2,f}(z) = \eta_f(z) = u_f^{-1}\left(u'(\eta(z))\right).$$

The first-order conditions for investment and savings are the same as with imperfect competition in financial markets.

Substituting for date 2 consumption into the budget constraint at date 1, the intertemporal budget constraint for agents of type i is:

$$c_{1i} + k_i + s_i + \sum_{z \in Z} q(z)\, c_{2i}(z) = e + \sum_{z \in Z} q(z)\left(y_i(z)\, k_i + R s_i\right). \tag{40}$$

Substituting the first-order conditions for k_i and s_i into (40), we arrive at:

$$c_{1,i} + \sum_{z \in Z} q(z)\, c_{2i}(z) = e. \tag{41}$$

Finally, substituting $\frac{c_{2i}(z)}{c_{1i}} = \eta(z)$ from (39) into (41), and recognizing $q(z) = \pi(z)\, u'(\eta(z))$, we find that:

$$c_{1i} = \frac{e}{1 + \sum_{z \in Z} \pi(z)\, u'(\eta(z))\, \eta(z)},$$

and therefore

$$c_{2i}(z) = \frac{\eta(z)\, e}{1 + \sum_{z \in Z} \pi(z)\, u'(\eta(z))\, \eta(z)},$$

Consequently, strategic agents consume in proportion to their initial endowments. Notice that this implies that:

$$\frac{\sum_{i=1}^{N} c_{2i}(z)}{\sum_{i=1}^{N} c_{1i}} = \eta(z). \tag{42}$$

Consider the type-symmetric case in which all agents have the same initial wealth e. Substituting the market clearing conditions at both dates into (42), and equating $\eta(z)$ with consumption growth in (38) and state prices in (39), we arrive at:

$$\Lambda^{CE}(z) = \pi(z)\, u'\left(\frac{\sum_{i=1}^{N} y_i(z)\, k + A^{CE}(z)}{N(e - s - k) + m_f\left(e - c_{f,1}^{CE}\right)}\right). \tag{43}$$

B Data Sources

In this Appendix, we discuss the data we use in Section 4. Our data come from three main sources. We measure risk-free rates using data on nominal rates on 1-year Treasury bills from the St. Louis Fed FRED database. Specifically, we use data series RIFSGFSY01NA (1-Year Treasury Bill Secondary Market Rate, Discount Basis, Percent, Annual, Not Seasonally Adjusted). This series shows values of 1.67% on 01/01/2002, and 0.6% on 01/01/2016. Figure 3 plots the data. Because there are missing values for the 1-year bill, we also plot 6-month bill rates in blue.

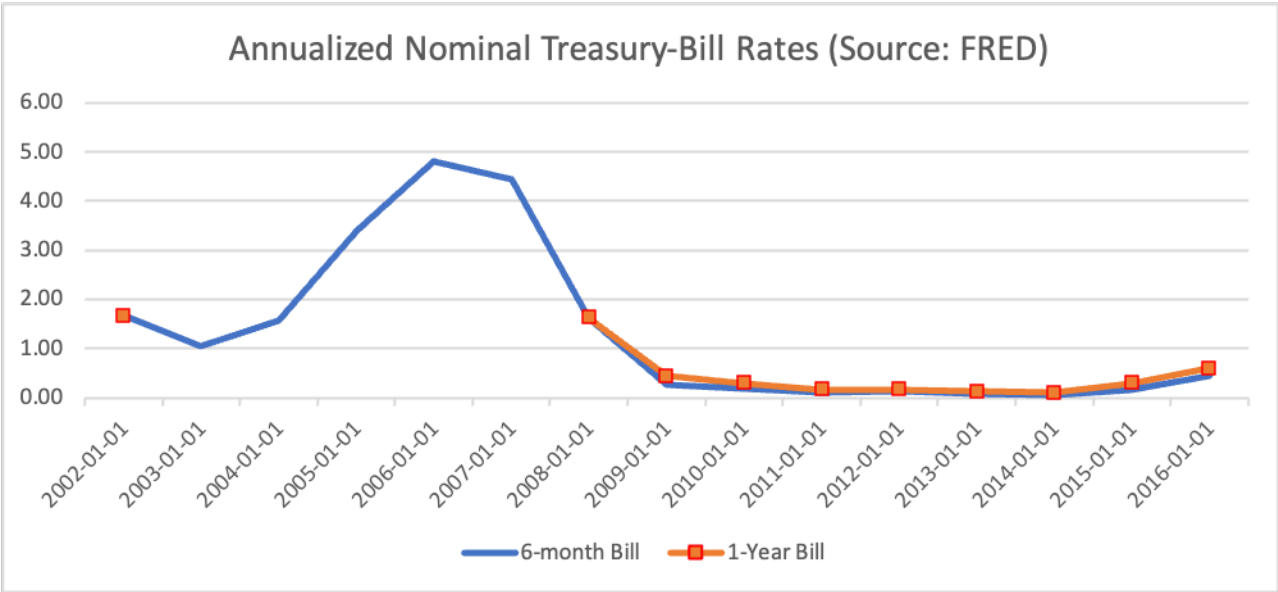

Figure 3: Interest rate data.

.

Data on corporate discount rates and weighted average cost of capital is from Gormsen and Huber (2022), who make data available at www.costofcapital.org. We use the raw average of annual rates. The data was retrieved on June 28, 2023. In our model, we refer to the "perceived cost of capital" of Gormsen and Huber (2022) as the working average cost of capital (WACC), and to the hurdle rate as the discount rate. Figure 4 plots the data series we use to calibrate our model.

We obtain data on firm-level dispersion using the interquartile range of firm-level Total Factor Productivity (TFP) from Figure 4, Panel b of Cunningham, Foster, Grim, Haltiwanger, Pabilonia, Stewart, and Wolf (2022). We thank the authors for sharing the underlying data. Our focus is on the years 2002 and 2016. Because of year-to-year vari-

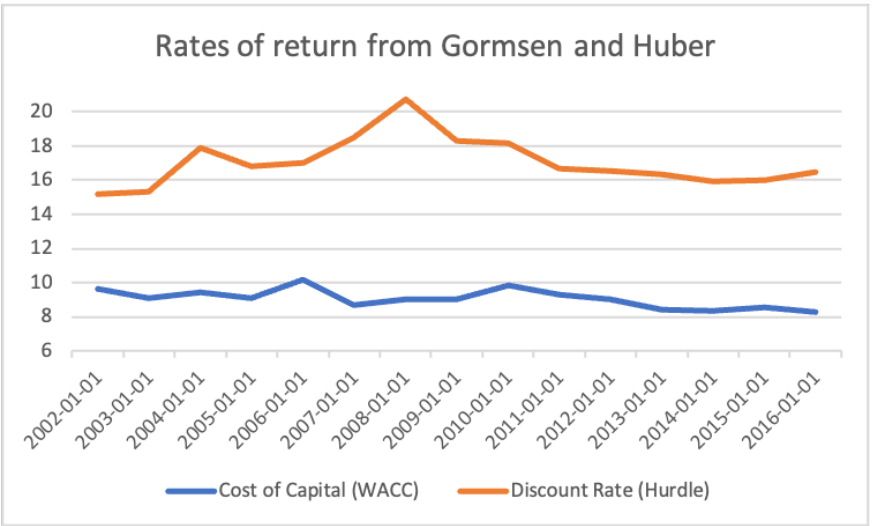

Figure 4: Data on cost of capital and discount rates.

.

ation in estimated TFP, comparing these two years only masks the underlying trend in dispersion over time. We therefore compute linear trends and calibrate to one half of the difference between the 75th percentile and the 25th percentile of this trend.

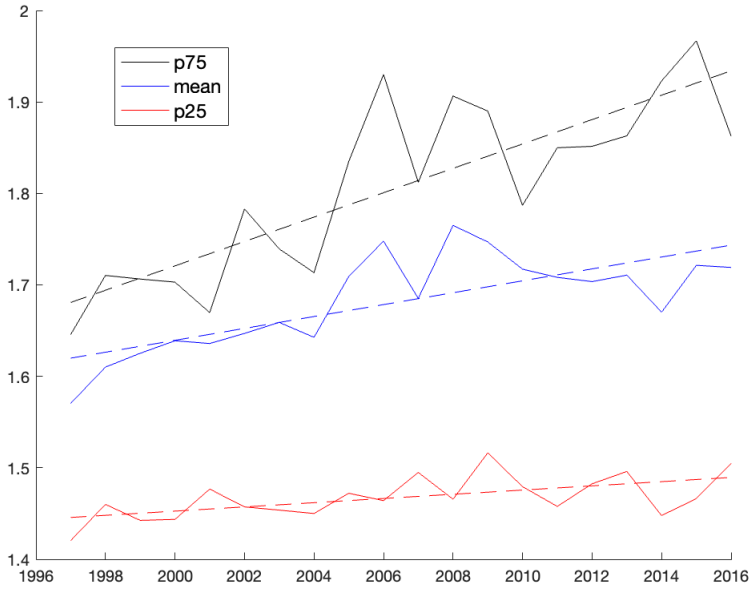

Figure 5: Interquartile range of firm-level TFP. Raw data is plotted in solid lines, linear trend is plotted in dashed lines. .

.

Online Appendix for Financial Market Concentration and Misallocation

Comparing Equilibrium-in-Demand-Schedules and Cournot-Walras Equilibrium

Daniel Neuhann and Michael Sockin

Our model of strategic trading in financial markets uses Cournot-Walras equilibrium as our equilibrium concept. This equilibrium concept differs from a long tradition following Kyle (1989), which focuses on Equilibrium in Demand Schedules (also known as double auctions). While both equilibrium concepts allow strategic traders to submit price-contingent demand schedules taking into account their impact on equilibrium prices, they have subtle differences that render each particularly suitable for some applications but not for others.

To understand these differences, we now present a canonical one-asset CARA-normal model that allows us to clarify commonalities and differences in the Equilibrium-in-Demand-Schedules and Cournot-Walras equilibrium concepts. We first solve for the competitive case as a benchmark, and then the Equilibrium-in-Demand-Schedules and Cournot-Walras approaches, respectively. We then discuss the outcomes of the two strategic equilibrium concepts.

Consider a two period model where all agents have CARA utility overall final wealth with parameter γ. A single risky asset is traded in period 1 and pays dividend X in period 2 where $X \sim \mathcal{N}(\bar{X}, \sigma_X^2)$. There is no discounting across periods and there is a riskless asset that pays zero interest in elastic supply.

There are N types of traders. Each type i consists $\frac{1}{\mu}$ agents with mass μ . Each agent of type i has initial endowment μz_i of shares of the asset so that their total endowment is z_i. The total endowment is consequently $Z = \sum_{i=1}^{N} z_i$.

An agent k of type i submits demand schedule $a_{k,i}(p)$ per its unit of mass to solve the following optimization problem:

$$u_{j,i} = \max_{a_j} E[-e^{-\gamma((z_k+a_k)(X-p(a_k))+p(a_k)z_k)}]$$

In what follows, we will conjecture and verify that the equilibrium asset price is normally distributed. In this case, the optimization program of agent k of type i can be expressed

as:

$$\max_{a_k} pz_k + (z_k + a_k)(\bar{X} - p) - \frac{\gamma}{2}(z_k + a_k)^2 \sigma_X^2.$$

There is also a competitive fringe. In accordance with the tradition following Kyle (1989), it acts as noise traders for now. They demand a random position $-U \sim \mathcal{N}(0, \sigma_U^2)$ net of the asset supply. For simplicity, U and X are independent. Market clearing imposes that

$$\sum_{i=1}^{N} a_i(p) = U. \tag{44}$$

Case 1: The Competitive Case

It is straightforward to show when all agents are price takers that

$$z_i + a_i = \frac{\bar{X} - p}{\gamma \sigma_X^2}.$$

Imposing market clearing (44), we recover the asset price

$$p = \bar{X} - \frac{\gamma \sigma_X^2}{N}(Z + U).$$

This completes our characterization of the competitive case.

Case 2: The Case with Equilibrium-in-Demand-Schedules

We now consider strategic behavior when the equilibrium concept is Equilibrium-in-Demand-Schedules. In this case, each strategic agent internalizes that they can influence the price p by shifting each other strategic agents' demand curves. Let us conjecture strategic agents of type i have a linear demand schedule:

$$a_i(p) = a_0 + a_p p - z_i. \tag{45}$$

Imposing market clearing (44), one has that:

$$\sum_{i=1}^{N} a_i(p) = U,$$

and by substituting with (45), strategic agent i's inverse demand curve is:

$$p = -\frac{\mu}{a_p}\frac{a_i}{N-\mu} - \frac{a_0}{a_p} + \frac{1}{a_p}\frac{Z+U}{N-\mu}.$$

This implies by rational expectations that strategic agent i's price impact is:

$$\frac{\partial p}{\partial a_i} = -\frac{1}{a_p}\frac{\mu}{N-\mu}. \tag{46}$$

The first-order condition for agent k of type i's optimal holdings is:

$$a_i = \frac{\bar{X} - p}{\mu\frac{\partial p}{\partial a_i} + \gamma\sigma_X^2} - z_i. \tag{47}$$

Consistency of the optimal holdings with the conjectured demand schedule implies:

$$a_p = -\frac{1}{\mu\frac{\partial p}{\partial a_i} + \gamma\sigma_X^2},$$

from which follows by substituting with (46):

$$\begin{aligned} a_p &= -\frac{1}{\gamma\sigma_X^2}\frac{N-\mu-\mu^2}{N-\mu}, \\ a_0 &= \frac{1}{\gamma\sigma_X^2}\frac{N-\mu-\mu^2}{N-\mu}\bar{X}, \end{aligned}$$

such that:

$$a_i(p) = \frac{N-\mu-\mu^2}{N-\mu}\frac{\bar{X}-p}{\gamma\sigma_X^2} - z_i. \tag{48}$$

The equilibrium price is then:

$$p = \bar{X} - \frac{N-\mu}{N-\mu-\mu^2}\frac{\gamma\sigma_X^2}{N}(Z+U), \tag{49}$$

and equilibrium price impact from (46) is:

$$\frac{\partial p}{\partial a_i} = \frac{\mu}{N-\mu-\mu^2}\gamma\sigma_X^2. \tag{50}$$

This completes our characterization of the equilibrium-in-demand-schedules case.

Case 3: The Case with Cournot-Walras Equilibrium

We now turn to Cournot-Walras Equilibrium. The key difference between the Equilibrium-in-Demand-Schedules approach and Cournot-Walras is whose demand curve strategic agents internalize. In the former, it is the functional form of the other strategic agents. In the latter, it is that of the competitive fringe. To this end, we now modify the competitive fringe to have CARA utility and an endowment U. This implies a demand a_f from the fringe similar to agents in the competitive case:

$$a_f = \frac{\bar{X} - p}{\gamma \sigma_X^2} - U.$$

Market clearing now imposes that

$$\sum_{i=1}^{N} a_i(p) + \frac{\bar{X} - p}{\gamma \sigma_X^2} = U. \tag{51}$$

This demand schedule of the competitive fringe implies by market clearing (51) that:

$$p = \bar{X} + \gamma \sigma_X^2 \sum_{i=1}^{N} a_i(p) - \gamma \sigma_X^2 U. \tag{52}$$

The demand of strategic agent k of type i again is:

$$a_i = \frac{\bar{X} - p}{\mu \frac{\partial p}{\partial a_i} + \gamma \sigma_X^2} - z_i. \tag{53}$$

Substituting with (55), i.e., $\frac{\partial p}{\partial a_i} = \gamma \sigma_X^2$, this implies:

$$a_i = \frac{1}{1+\mu} \frac{\bar{X} - p}{\gamma \sigma_X^2} - z_i. \tag{54}$$

Substituting (54) into (52), the equilibrium price is:

$$p = \bar{X} - \frac{1+\mu}{N+1+\mu} \gamma \sigma_X^2 (Z + U), \tag{55}$$

and the equilibrium price impact is again:

$$\frac{\partial p}{\partial a_i} = \gamma \sigma_X^2 \tag{56}$$

Comparing Equilibrium Concepts

We now discuss the key similarities and differences between the two strategic equilibrium concepts. We then highlight several advantages of each concept to conclude the note.

Similarities

The most important similarity is that strategic agents face identical portfolio trade-offs in both equilibrium concepts. In particular, the optimization problem of the strategic agents are the same in the precise sense that (48) and (54) are identical. This is because strategic agents in both the Equilibrium-in-Demand-Schedules and Cournot-Walras concepts submit price-contingent demand schedules that balance the benefit of a better portfolio allocation against the cost of price impact. In both concepts, portfolio choices a_i therefore encode the same trade-offs. Moreover, the degree to which agents extract rents by manipulating prices based on the illiquidity of financial markets, measured by the price impact term $\frac{\partial p}{\partial a_i}$.

Differences

The key difference between the two equilibrium concepts is the precise origin of the price impact term. This difference is driven by differences in the identity of the agents whose demand strategic agents believe they can manipulate. In the Equilibrium-in-Demand-Schedules approach, it is the demand curves of other strategic agents (i.e., the price input to their demand schedules). This is appealing because such models often feature exogenous noise traders, who are insensitive to the price, and allows for a rich variety of strategic interactions among large agents. However, it is perhaps less well-suited to settings where agents can trade many assets with varies counterparties in different markets, and assets may be fungible with respect to the risk exposure they deliver. In such markets, inverse demand functions for individual assets and certain participants may be difficult to forecast, and they may not be sufficient statistics for price impact. In such settings, there

is an advantage to directly deriving price impact in terms of the change in the distribution of risk induced by a trade.

This is what occurs in the Cournot-Walras approach, where price impact is derived from the demand curve of a price-taking competitive fringe. That is, the sufficient statistic for price impact is the demand function of the relatively unsophisticated investors (the competitive fringe), which is a simpler object to characterize and is valid for any asset with any payoffs as long as one knows the payoffs. Accordingly, the equilibrium price impact term in our example model is (56), which reflects only the incremental change in risk borne by the market $\gamma\sigma_X^2$. This is because each strategic agent takes other strategic agents' demands as given, so that the competitive fringe absorbs this incremental demand.

As we discuss below, this approach has the advantage that it can be flexibly adapted to any market structure and allows for general equilibrium effects as well investment. Moreover, the approach is rooted in a rich tradition studying product market competition among Cournot oligopolists. As such, the equilibrium concept has well-known desirable properties.

Several Advantages of Cournot-Walras

Although the Equilibrium-in-Demand-Schedules concept is an appealing concept, it has several drawbacks that Cournot-Walras does not. First, is that solving the equilibrium becomes extremely difficult outside of the CARA-normal setting. This is because a strategic agent must correctly forecast the demand schedules of other strategic agents to form beliefs about her equilibrium price impact, which is a fixed-point object. Outside of the CARA-normal setting, this inverse demand curve need neither be linear nor unique; in fact, only in the CARA-normal case is the unique inverse demand curve linear. Consequently, there can be multiple equilibria in which demand curves take non-obvious functional forms that are highly dependent on the setting being analyzed. This has limited most applications of the Equilibrium-in-Demand-Schedules approach to the CARA-normal paradigm where the payoffs of assets must be exogenously specified and cannot be determined in equilibrium.

The Cournot-Walras approach does not have this drawback. As long as one can specify the preferences of a competitive fringe, price impact is uniquely determined by

the fringe's preferences for any payoff structure. This not only avoids the issue of conjecturing and verifying an inverse demand curve, but also shuts down strategic uncertainty among large agents as to which equilibrium price impact function will prevail (i.e., among many potential price impact functions with other strategic equilibrium concepts). In this sense, even in the limit that the fringe becomes arbitrarily small, and market clearing becomes effectively internal among strategic agents, we have a well-defined selection mechanism for identifying an unique price impact function.

This flexibility allows the Cournot-Walras approach to be applied in quite general settings with quite general preferences. This may be more appropriate for examining the market-wide consequences of market concentration, in which it is not obvious that large agents face regular trading partners over time. In such a context, the competitive fringe has a natural interpretation as households and other retail traders. It also does not restrict the market structures that one can study to a linear asset of normal random variables, which is a very incomplete market structure.

Several Advantages of Equilibrium-in-Demand-Schedules

The Equilibrium-in-Demand-Schedules concept offers several advantages over Cournot-Walras. First, is that it allows for more complex forms of strategic interaction among large agents. Such nuanced interactions can give rise to interesting forms of equilibrium behavior. This is often why papers in this literature focus on this rich object. In addition, such an approach may be more appropriate in thin markets that are dominated by a few large market participants that do not vary much over time. In this situation, it makes sense that large agents would understand and internalize how each other will react to their trading strategies. Finally, one does not need to take a stand on the preferences of a competitive fringe that pins down equilibrium price impact.

That equilibrium price impact is determined by the trading behavior of the strategic agents is an appealing advantage over the Cournot-Walras concept, in which price impact is less nuanced. Price impact in the Cournot-Walras setting is instead derived from the marginal utility of a competitive fringe after market clearing is imposed. This anchors the strategic interactions among large agents to the choice of the fringe's preferences and optimization problem, which may be an undesirable degree of flexibility.

Conclusion

As we have demonstrated in a canonical CARA-normal setting, the Equilibrium-in-Demand-Schedules and Cournot-Walras approaches are not that different. Both lead to similar trade-offs for strategic agents when choosing their optimal portfolios. What is different is how we arrive at equilibrium price impact. With the Equilibrium-in-Demand-Schedules approach, this price impact is based on how each other strategic will react to a marginal change in demand. With Cournot-Walras, price impact is determined by how a competitive fringe will react to a marginal change in demand.

Both concepts consequently have their advantages and drawbacks, and which is more appropriate likely depends on the setting and given application. Important for many applications, such as examining how market concentration interacts with risk sharing and wealth effects, is allowing for curvature in the marginal utilities among strategic agents. This gives rise to asymmetries in the marginal benefits / costs of trading an additional unit depending on whether the strategic agent is a buyer or a seller, or wealthier or poorer. In CARA-normal settings, marginal utility is linear, and there is therefore no asymmetry between buying or selling an additional unit, although differences in risk aversions can lead to differences in marginal benefits / costs between strategic agents. In such situations, the limitations of relegating equilibrium price impact to effectively a primitive for strategic agents (since we can specify the competitive fringe's utility however we want) may be outweighed by the ability to examine these important issues.