Portfolio Regulation of Large Non-Bank Financial Institutions*

Daniel Neuhann[†] Mahyar Sefidgaran[‡] Michael Sockin[§]

November 2023

Abstract

We study the role of regulatory mandates, or investment restrictions, in shaping the distribution of risk taking and market power among large non-bank financial institutions. Institutions trade to hedge risks, but market concentration leads to constrained-inefficient risk sharing even when all institutions can access complete financial markets. Imposing mandates on a subset of investors redistributes market power and can further hamper risk sharing. Optimal market-wide mandates, in contrast, can improve risk sharing and welfare through general equilibrium forces, and we characterize their properties. Our findings suggest that broad coverage is critical for implementing effective portfolio constraints on non-bank financial institutions.

*This paper was previously titled "Market Power and Missing Markets." We thank Ana Babus, Bradyn Breon-Drish, Eduardo Dávila, William Fuchs, Kinda Hachem, Ben Hébert, Giorgia Piacentino and a number of seminar audiences for helpful comments.

[†]University of Texas at Austin. daniel.neuhann@mccombs.utexas.edu. (Corresponding author).

[‡]University of Texas at Austin. mahyar.sefidgaran@mccombs.utexas.edu

[§]University of Texas at Austin. michael.sockin@mccombs.utexas.edu

1 Introduction

An ongoing challenge in modern financial markets is how to regulate large non-bank financial institutions (NFBIs) such as pension funds and insurance companies. These institutions invest alongside other intermediaries, such as banks, in a number of systemically important financial markets, including relatively illiquid asset classes like corporate bonds, real estate, private equity, or interest rate swaps. As such, the stability of NFBIs has become an important component of the stability of the financial system overall. Most recently, this became apparent when a number of U.K. pension funds suffered losses that threatened their solvency after the Bank of England raised interest rates in September 2022, forcing regulators to intervene in order to forestall a broader crisis event.

How can such events be prevented? Historically, banks have been the central focus of financial regulation. Accordingly, much of the literature on optimal regulation considers settings in which excessive leverage, subsidized deposit insurance, or moral hazard in origination can trigger excessive risk taking. While these are critical concerns for banks, they are less central to the business model of NBFIs, who must worry more about the management of payout and duration risk, and their price impact in thin financial markets (e.g., Khetan, Neamtu, and Sen (2023)). It is consequently unclear whether existing regulatory frameworks apply to NBFIs, or if new tools are needed. In this paper, we develop and assess some of these new tools for NBFI regulation.

In particular, a common regulatory instrument that has been used to foster sound risk management at large NFBIs is restrictions on what assets they can trade.[1] However, there is no consensus on how to optimally design such restrictions. Some jurisdictions allow NFBIs to trade complex securities such as interest rate and foreign exchange derivatives. Others allow them to trade only in liquid and standardized securities such as stocks and bonds. For instance, while the EU and South Africa prohibit pension funds from trading derivatives, the US and UK encourage investment in alternative asset classes.[2] Similarly, The E.U. prohibits liability-driven investment funds from using derivatives and

[1]Such restrictions can be imposed privately through investment mandates or publicly through portfolio regulation that targets either a subset of or all non-bank financial institutions.

[2]New York recently passed a bill raising the investment cap for pension funds on alternative assets, while the UK relaxed its stringent default fund charge cap for defined contribution funds to allow for performance fees that could support investment in alternative assets. OECD (2015) and OECD (2021) offer international surveys of insurance company and pension fund investment regulation.

there are now calls from U.K. lawmakers for the Bank of England to regulate the use of derivatives by pension funds.[3]

There are sound arguments in favor of more permissive mandates. For instance, classical theory suggests that unfettered access to state-contingent claims should improve risk sharing. However, empirical evidence suggests that this is not always true. For example, pension funds and insurance companies appear to under-insure or even amplify their balance sheet exposures using derivative markets (e.g., Pinter and Walker (2023)), in part because demand pressure from some market participants makes it more difficult for these institutions to trade in illiquid markets (Khetan, Neamtu, and Sen (2023)).[4] This has rekindled interest in the optimal determination of investment mandates and portfolio regulation. Given the role of demand pressure in determining asset allocations, this question cannot be disentangled from its interaction with market structure and liquidity.

Understanding the effects of investment mandates and portfolio regulation in the presence of price impact is challenging for several reasons. First, institutional investors may be able to sidestep restrictions on specific assets if they can construct replicating portfolios that deliver the same underlying cash flows. As such, regulating access to one instrument may simply push trading activities into another market, with either neutral or unintended consequences. For instance, Jansen (2021) shows that changes in regulation shifted Dutch pension fund and insurance company demand from short-term to longer-term twenty-year Dutch government bonds, steepening the yield curve and reallocating short-term bonds to banks. Second, large investors must behave strategically when trading in illiquid asset markets, and they interact with other strategic investors across multiple markets simultaneously. Finally, market liquidity is itself influenced by the set of restrictions imposed on large investors, and this in turn impacts their optimal design. Addressing these issues requires a general equilibrium model of strategic trading in which market liquidity is endogenously determined and investment universes are flexible. Our contribution is to provide such a framework and to make progress in characterizing the positive and normative implications of imposing restrictions on what large

[3]See https://www.reuters.com/business/finance/lawmakers-call-bank-england-role-pensions-after-ldi-debacle-2023-02-07/.

[4]The risk of margin calls on interest rate derivatives further amplified the exposure of U.K. pension funds to interest rate hikes in September 2022. Jansen, Klinger, Ranaldo, and Duijm (2023) shows that this is true for Dutch pension funds as well.

institutional investors can trade.

We study the role of investment mandates and portfolio regulation in a two-period general equilibrium economy with arbitrary asset span. We model markets as complete, but allow a regulator to determine arbitrary restrictions on what agents can trade. As a consequence, mandates can render markets functionally incomplete. There are gains from trade because investors have different income processes and/or risk preferences, and manage these risks in financial markets. There is a discrete number of large non-bank financial institutions who internalize by virtue of the size of their trading needs that their trades move prices, and a mass of price-taking traders (the competitive fringe). The model is tractable: prices and price impact functions are equal to the first and second derivatives of the competitive fringe's utility, asset prices satisfy no arbitrage for any mandates, and we can vary market power without affecting the aggregate consumption frontier. Moreover, because of no arbitrage, equilibrium allocations are invariant to the choice of market structure among those with equivalent asset spans. As such, introducing redundant assets does not affect prices or allocations, which is important because it ensures that our model is robust to investor attempts to sidestep mandates by trading other portfolios of assets.

We begin with a benchmark in which there are no restrictions on what NFBIs can trade. Absent mandates or regulation, privately-optimal distortions can be summarized using state-contingent wedges in the first-order condition relating the marginal valuation of consumption in a given state (i.e. the state price) to the price of the Arrow security referencing that state. As is well-appreciated, a seller reduces her supply of a security to inflate its price above her state price, while a buyer reduces her demand to depress it below. We add to this insight by showing that these distortions are most severe when *prices* are high, because this reflects high equilibrium marginal valuations. Importantly, they are also exacerbated by an illiquidity externality that arises because price impact is nonlinear and large investors do not internalize how changes in their asset positions indirectly affect each other's incentives to exert market power.

Although conceptually appealing, formulating inefficiencies as state-contingent wedges at the investor-level provides only limited insight into the general equilibrium distortions that they entail. This makes it difficult to design appropriate market-wide regulatory mandates. To overcome this issue, we map state-contingent wedges into market-wide risk-sharing arrangements by constructing a map between the equilibrium of our model with complete markets and an *equivalent counterfactual economy with competitive trading but incomplete markets*. Such a mapping is feasible because market power leads to misaligned marginal valuations (i.e., state prices) similar to what occurs in models with incomplete markets. The particular form of incompleteness of the counterfactual (i.e. which markets appear to be missing) then provides a clean measure of the directions in which trade is distorted by strategic trading in the aggregate.

More specifically, we decompose the Arrow security markets into a *traded* and a *rationed* asset span. This decomposition reveals a taxonomy of the types of risks that are shared well in equilibrium (the traded asset span), and the types of risks that are shared insufficiently because of market power (the rationed asset span). Using two canonical settings (one with ex-ante symmetric agents subject to diversifiable risk, the other with heterogeneous preferences), we show that the rationed asset span is directly related to the unrealized risk-sharing opportunities, and that the degree of distortions are growing in the underlying gains from trade. Large non-bank financial institutions consequently have incentive to distort their portfolios to balance extracting surplus from financial markets at the expense of risk management. As we show below, the rationed asset span provides direct insights into optimal mandates in these canonical settings. Our mapping into a counterfactual asset span may also be of independent methodological interest as a tool for characterizing the aggregate distortions from frictions that hamper risk sharing.

We then introduce investment mandates and portfolio regulation, which represent constraints on what institutional investors can trade. Our first positive insight is that investment mandates and narrow portfolio regulation that apply only to what a subset of large institutional investors can trade reallocates market power to those that are unconstrained. Investors with mandates cannot implement their optimal distortions state-by-state, while those without mandates more aggressively ration their constrained trading partners. In practice, this means that mandates on regulated institutions such as pension funds may reallocate pricing power to investors with fewer constraints, such as hedge funds.

This improves market liquidity in markets in which investors with mandates are sellers and worsens those in which they are buyers, which has an ambiguous impact on welfare. This insight is particularly relevant because pension funds and insurance companies, which are highly regulated, often trade against asset managers like hedge funds (Khetan, Neamtu, and Sen (2023), Pinter and Walker (2023)) that are less regulated and able to strategically exploit the constraints imposed by regulation.

Next, we show the striking result that suitably chosen market-wide portfolio regulation, defined as restrictions on what *all* market participants can trade, can generically improve both liquidity and welfare, despite imposing direct costs on trading. This is because such policy can mute the illiquidity externality that stems from the non-linearity of price impact and gives rise to deadweight losses. More broadly, this result can be interpreted as showing that, given price impact, there always exist incomplete market structures that deliver better risk sharing than complete markets. The main intuition for this result is that relatively well-capitalized sellers of assets tend to distort their trades more than relatively poor buyers (after all, buyers buy because they have a high willingness to pay). Suitably chosen regulation improves overall welfare and liquidity because it redistributes consumption away from sellers and towards buyers, thereby reducing the deadweight losses from the illiquidity externality.

Although a fully general analysis of optimal regulation is beyond the scope of our paper, we provide insight into the optimal design of mandates using two canonical settings, one with ex-ante symmetric agents sharing diversifiable risk and the other with heterogeneous preferences and aggregate risk. We show that the optimal regulatory mandates in these two cases are precisely the assets identified as those rationed by our counterfactual economy with competitive trading. In particular, the optimal asset span consists of a single asset that serves to reallocate the most important risks but does not permit additional trading of other risk exposures. This shows that narrow asset spans can be advantageous if they are aligned with the fundamental direction of gains from trade.

Overall, our key policy implication is that regulators need to take a broad view toward the risk management of large non-bank financial institutions, and cannot impose portfolio restrictions on only a subset in the presence of market concentration. Regulators need to regulate all large institutional investors, and not just pension funds or insurance companies, to curtail their rent-seeking behavior and improve their risk management. In practice, insurance companies in the United States are regulated at the state-level, pension funds by the Department of Labor, and asset managers by the Securities and Exchange Commission. Our analysis highlights that in the new post-financial crisis financial market environment, regulators of NFBIs must work together in the design of optimal regulation.

Related literature. Our analysis contributes to a growing literature studying the behavior and regulation of large financial institutions. Egan, Hortacsu, and Matvos (2017) examines the role of regulation in promoting financial stability among large banks that hold uninsured deposits, while Buchak, Matvos, Piskorski, and Seru (2018) study how regulation provided an advantage to Fintech lenders over traditional banks. Similar to Buchak, Matvos, Piskorski, and Seru (2022), who study how capital requirements changes and quantitative easing impact competition among traditional and shadow banks, we emphasize the importance of accounting for the industrial organization among large NFBIs when designing regulation. Basak and Pavlova (2013) illustrate how institutional investors distort asset prices when their preferences take into account the performance of a benchmark. Hachem and Song (2021) study liquidity booms in imperfectly competitive interbank markets. Gabaix and Koijen (2021) show how financial markets become inelastic when institutional investors have limited ability to reallocate capital across markets. Koijen and Yogo (2019) and Haddad, Huebner, and Loualiche (2021) analyze how large investors choose their portfolios using a reduced-form demand-system approach. Different from these studies, we analyze how the risk management and rent-extraction behavior of large institutional investors interacts with the assets that they can trade.

Our paper is also related to the literature on endogenous market incompleteness in which risk sharing is limited by constraints, such as limited commitment (e.g. Kehoe and Levine (1993), Alvarez and Jermann (2000), Hellwig and Lorenzoni (2009)). In these models, complete markets continue to allow the maximum feasible gains from trade to be realized and state prices are always fully aligned. As a result, they do not permit a formal equivalence with a restricted asset span or an analogy to rationed asset markets. Our restricted asset span equivalence result invites comparisons to models of risk sharing in incomplete markets with perfect competition. This literature shows competitive markets are generically constrained inefficient when there are multiple goods (e.g., Hart (1975)), incomplete information (e.g., Greenwald and Stiglitz (1986)), or borrowing constraints (e.g., Dávila and Korinek (2018), Bocola and Lorenzoni (2020)). We show that when large agents have price impact, restricting what they can trade can improve welfare by attenuating the distortionary impact of market power.

Our equilibrium concept is Cournot-Walras in the tradition of Gabszewicz and

Vial (1972). The main benefit of this concept is that we can incorporate rich heterogeneity in preferences and income risk, arbitrary asset spans, asymmetric strategies, and nonlinear price impact, all of which are important in our finding that optimal distortions are highly state-contingent. Using a Cournot framework, Basak (1997) explores how having a monopolistic non-price-taking agent acts as a market leader when sharing risk with price-taking agents in an Arrow-Debreu economy. Rahi and Zigrand (2009) illustrate the inefficiencies in the incentives of large traders to arbitrage across segmented markets. Eisenbach and Phelan (2022) study fire sales externalities while Kacperczyk, Nosal, and Sundaresan (forthcoming) investigate asset price informativeness. Neuhann and Sockin (2023) uses a related model to study implications of strategic trading for real investment, but assumes markets are complete. Different from these studies, we examine how risk sharing among oligopolistic investors interacts with what assets they can trade.

A related approach based on Kyle (1989) instead studies equilibrium-in-demand-schedules. This concept allows for richer strategic interactions among large traders at the cost of stronger assumptions on preferences and payoffs (i.e., symmetry, CARA-normal settings) for tractability. The most suitable concept is context-specific, and Cournot-Walras is particularly useful in our setting because it preserves the basic strategic forces of market power while allowing us to vary the asset span without affecting the underlying gains from trade.[5] Building on the Kyle (1989) approach, Malamud and Rostek (2017) and Rostek and Yoon (2021) show that introducing redundant assets or restricting trading partners through decentralization in over-the-counter markets can improve risk sharing by redistributing price impact across traders. We highlight the complementary result that *restricting the asset span* (rather than trading partners) may improve welfare. To our knowledge, we are also the first to derive a taxonomy distinguishing assets valuable for risk sharing from those that incrementally facilitate rent extraction, and to illustrate how this can inform financial market regulation.

In this context, our findings are related to the literature on financial innovation and market design (e.g., Demange and Laroque (1995), Pesendorfer (1995)). Athanasoulis and Shiller (2000) show a social planner in a CARA-normal setting will first open asset markets most aligned with competitive agents' endowments. Previous work has established closing markets may be optimal when financial markets are constrained inefficient, such

[5] We provide an in-depth discussion of the two equilibrium concepts in Neuhann and Sockin (2023).

as with multiple goods (e.g., Cass and Citanna (1998), Elul (1995)) asymmetric information (e.g., Marin and Rahi (2000)), or heterogeneous beliefs (e.g., Blume, Cogley, Easley, Sargent, and Tsyrennikov (2018)). Carvajal, Rostek, and Weretka (2012) show how profit maximizing security design by competitive agents may involve leaving markets incomplete. Babus and Hachem (2020) and Babus and Hachem (2021) consider how private incentives to design risky securities depend on the competitiveness of the demand side, while Babus and Parlatore (2021) explore market fragmentation with endogenous liquidity. In contrast to this literature, we find that trading in unrestricted, complete markets *increases* the scope for privately optimal (but socially inefficient) rent extraction with imperfect competition.

2 Model

We study an endowment economy with a continuum of price-taking agents (the competitive fringe) and a finite number of large agents with market power. The fringe ensures that asset prices satisfies no arbitrage and can be solved analytically. Markets are, in principle, complete but we allow for any arbitrary financial market structure to examine a rich set of restrictions, i.e., private mandates or regulatory constraints, on what large agents can trade. In this precise sense, our model is a generalization of the one used in Neuhann and Sockin (2023). In Section 3, we focus on the baseline case when there are no restrictions and all risks can be traded to clarify how market power relates to risk sharing incentives. In Section 4, we consider restricted asset spans to show how restricting which assets that large institutional investors can trade can improve risk sharing under market power.

Agents, endowments, and preferences. There are two dates, $t = \{1,2\}$ and a single numeraire good. Uncertainty is modeled as a discrete set of states of the world $\mathcal{Z}$ with cardinality $Z = |\mathcal{Z}|$. State $z \in \mathcal{Z}$ is realized at date 2 with probability $\pi(z) \in (0,1)$.

There is a price-taking competitive fringe of mass m_f and a finite number of strategic agents who internalize their actions affect equilibrium prices. There are $N > 1$ *types* of strategic agents, indexed by $i \in \{1, 2 \ldots, N\}$. We assume $N \leq Z$ so that there are weakly fewer agent types than states of the world, although this assumption is only required for Propositions 3 and 4.

Agent k of type i receives a state-contingent income $y_i(z)$ in state z, and an initial endowment of $w_{k,i}$ at date 1. These agents represent large financial institutions, such as insurance companies and pension funds or their strategic counterparties such as hedge funds, and their incomes the net state-contingent payoffs from premiums less payouts to insurees or defined benefit pensioners. Since we work in a complete-markets setting with essentially unrestricted income processes, the model can also be used to capture interest risk through a suitable redefinition of the states of the world. The fringe, which represents households and retail investors, receives state-contingent income $y_f(z)$ and initial wealth w_f. Let w_i be the total initial wealth of agents of type i, i.e., $w_i = \sum_k w_{k,i}$.

Define $\mu_i = \frac{w_{k,i}}{w_i}$ to be the fraction of initial wealth held by each agent of type i, and we assume μ is the same across all agent types, i.e., $\mu_i = \mu$ for all i. This has the interpretation of there being $\frac{1}{\mu}$ *symmetric* agents of each type, each with mass μ. For an individual agent k of type i, this μ parameterizes her market power because it represents the fraction that she has of the total wealth of type i, w_i, to manipulate prices in financial markets. In the aggregate, μ represents what we refer to as *market concentration*, and a higher μ reflects a higher degree of concentration. The competitive equilibrium corresponds to the special case in which $\mu = 0$. The total mass of strategic agents is N, and we focus on the case in which all agents within a type follow symmetric strategies.

All agents have homothetic type-specific utility concave functions $u_{i,t}(\cdot)$ over consumption $c_{i,t}$ at date t. Risk aversion captures the notion that even large financial institutions can exhibit risk aversion under a variety of frictions, such as capital and risk management constraints. The competitive fringe has linear utility over initial consumption, $u_{f,1}(c_{f,1}) = c_{f,1}$, and concave utility $u_f(\cdot)$ over date 2 consumption (i.e., quasi-linear preferences). Both $u_{i,t}(\cdot)$ and $u_f(\cdot)$ are $\mathcal{C}^2$, strictly increasing, and do not have linear marginal utility for all i. Quasi-linear preferences for the fringe are convenient because the pricing functional depends only on the fringe's marginal utility at $t = 2$, but the assumption is not essential.

Financial Markets. Markets are, in principle, complete in that all risks are tradeable in the absence of mandates or regulation on large agents. Introducing portfolio constraints on all large agents through regulation, however, effectively restricts the asset span they can trade. To capture this, we allow for an arbitrary asset span. Financial markets consequently consist of $J \leq Z$ securities with bounded payoffs $\{x_j(z)\}_{j=1}^{J}$. A market structure is indexed by a matrix $X \in \mathbb{R}^{J\times Z}$ such that $\vec{x}(z) = X\vec{\delta}(z)$, where $\vec{\delta}(z)$ is the $Z \times 1$ vector whose z^{th} entry is 1 and 0 otherwise. Trading takes place at date 1 and assets pay out at date 2.

Definition 1 *A market structure X is complete if $rank(X) = Z$ and restricted if $rank(X) < Z$.*

Because all $1/\mu$ agents of a given type are symmetric, we study equilibria that are symmetric within each type. We denote type i's asset position in security j by $a_i^j \in \mathbb{R}$, where $a_i^j < 0$ denotes sales. Let $\mathcal{K}_i$ denote the set of strategic agents of type i. It is understood an individual agent k of type i holds μa_i^j units of the asset, so the total holdings of all agents of type i is equal to a_i^j,

$$\sum_{k \in \mathcal{K}_i} a_i^j \mu = a_i^j. \tag{1}$$

The set of market clearing conditions can then be stated concisely as

$$\sum_{i=1}^{N} a_i^j + m_f a_f^j = 0 \qquad \text{for all } j. \tag{2}$$

Because equilibrium prices depend on asset positions, we denote the market-clearing pricing functional by $Q_j(\vec{a}^j)$, where $\vec{a}^j = [a_1^j, a_2^j, \ldots a_N^j]$ is the vector of agents' holdings of asset j. The *price impact* of an agent of type i for asset j is the marginal change in the equilibrium price of asset j given a marginal change in the agent's asset position, holding fixed other large agents' positions. We denote the price impact of the *representative* agent of type i by $\tilde{Q}_i(\vec{a}^j)$. By Equation (1), the change in quantities induced by a change in the portfolio position of an *individual* agent of type i depends on her mass μ; as such, the price impact of an individual agent of type i is $\mu\tilde{Q}_{i,j}(\vec{a}^j)$.

Equilibrium Concept. Our equilibrium concept is Cournot-Walras Equilibrium. In this concept, strategic agents submit prince-contingent demand schedules taking as given the each other's demand. Neuhann and Sockin (2023) provides a longer discussion of this concept, including similarities and differences with Kyle (1989).

At time 1, strategic agent i allocates endowment w_i between immediate consumption and asset purchases. At date 2, he consumes his income and asset payoffs. Strategic agents takes each other's asset demands as given when making portfolio choices, but they internalize they influence the competitive fringe's consumption.

The decision problem of the representative strategic agent of type i is

$$U_i = \max_{\left\{c_{i1},\{a_i^j\}_{j=1}^J\right\}} \quad u_{i,1}\left(c_{i1}\right) + \sum_{z\in\mathcal{Z}} \pi\left(z\right) u_{i,2}\left(c_{i2}\left(z\right)\right) \tag{3}$$
$$\text{s.t.} \quad \mu c_{i1} = \mu w_i - \mu \sum_j Q_j(\vec{a}^j) a_i^j,$$
$$\mu c_{i2}\left(z\right) = \mu y_i\left(z\right) + \mu \sum_j x_j(z) a_i^j,$$

We follow this convention for defining the controls of strategic agents recognizing the consumption of the representative strategic agent of type i is actually μc_{i1} and $\mu c_{i2}\left(z\right)$ at dates 1 and 2, respectively, and similarly with optimal asset holdings, μa_i^j. This is because under homothetic preferences optimal policies are invariant to μ.

The decision problem of the competitive fringe is

$$U_f = \max_{\left\{c_{f1},\{a_f^j\}_{j=1}^J\right\}} \quad c_{f1} + \sum_{z\in\mathcal{Z}} \pi\left(z\right) u_{f,2}\left(c_{f2}\left(z\right)\right) \tag{4}$$
$$\text{s.t.} \quad c_{f1} = w_f - \sum_j Q_j a_f^j,$$
$$c_{f2}\left(z\right) = y_f\left(z\right) + \sum_j x_j(z) a_f^j,$$

The key difference between the two problems is that strategic agents internalize that prices are a function of theirs and each other's quantities, while the competitive fringe treats prices as a constant. We call the equilibrium with price impact a *market equilibrium*.

Definition 2 (Market equilibrium) *A market equilibrium is a Cournot-Walras Equilibrium consisting of strategy profiles* $\sigma_i = \left(c_{i1}, \{a_i^j(z)\}_{j=1}^J\right)$ *for each representative agent of type i and* $\sigma_f = \left(c_{f1}, \{a_f^j(z)\}_{j=1}^J\right)$ *for the competitive fringe, pricing functions* $Q_j(\vec{a}^j)$ *and associated price impact functions* $\tilde{Q}_{i,j}(\vec{a}^j)$ *for each asset j such that:*

1. *Policy* σ_i *solves Decision Problem* (3) *for each i given* σ_{-i} *and the set of pricing functions.*

2. *Each market clears with zero excess demand according to* (2).

3. *Price impact functions are consistent with pricing functions for all assets.*

4. *All agents have rational expectations with respect to their equilibrium price impact.*

Strategic interaction in our model is intermediated by the competitive fringe. Although 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 *indirectly* affects how another strategic agent type trades by altering the price (and price impact) that agent type faces. In addition, because there is a unique mapping between the fringe's marginal utility and a strategic agent's demand, the equilibrium pricing function is unique. This is in contrast to the rich equilibrium-in-demand-schedules approach of Kyle (1989) that allows for more general pricing functions to study different cross-agent and cross-market price impact protocols within the same model. In those settings, there can be multiple equilibria indexed by very different price impact functions. Because our focus is on how asset spans affect risk sharing among strategic agents, we choose a concept in which price determination is intuitive and unique, given economic primitives, at the cost of abstracting from more complex strategic concerns.

To characterize the optimal asset of strategic agents, we define the state prices for agent i as the expected marginal rate of substitution between states z and time 1, i.e.

$$\Lambda_i(z) \equiv \frac{\pi(z)u'_{i,2}(c_{i2}(z))}{u'_{i,1}(c_{i1})},$$

where $f'(\cdot)$ is the derivative of $f(\cdot)$. In addition, let $\bar{\Lambda}_i$ be the vector of agent i's state prices and $\bar{a}_i$ her $J \times 1$ vector of asset demands. This allows us to compactly describe the following key properties of the equilibrium. Consistent with the notion of a Walras equilibrium, strategic agents submit price-contingent demand schedules along with the competitive fringe to a Walrasian auctioneer who sets the prices that clear markets. Importantly, because the competitive fringe enforces no arbitrage in financial markets, consumption allocations are invariant to the set of securities agents trade provided they have equivalent asset spans.

Proposition 1 *There exists a market equilibrium in which:*

1. *Securities prices satisfy the Law of One Price and are given by* $Q_j(\vec{a}^j) = q_j(\vec{a}^j)$ *where*

$$q_j(\vec{a}^j) = \sum_{z \in \mathcal{Z}} x_j(z)\Lambda_f(z) \quad \textit{and} \quad \Lambda_f(z) = \pi(z)u'_{f,2}\left(c_{f2}(z)\right).$$

The fringe's state price $\Lambda_f(z)$ is decreasing in consumption $c_{f2}(z)$ and, by market-clearing,

$$c_{f2}(z) = y_f(z) - \sum_j x_j(z) \frac{1}{m_f} \sum_{i=1}^{N} a_i^j.$$

2. *The price impact matrix of the representative agent of a given strategic agent type is symmetric across $i \in \{1, ..., N\}$,*

$$\tilde{Q}(\vec{a}^j) = \frac{\mu}{m_f} X \Gamma X' \quad \textit{where} \quad \Gamma_{j,z} = -\Lambda_f'(z) \mathbf{1}_{\{j=z\}} \geq 0;$$

3. *The optimal asset portfolio of type $i \in \{1., ..., N\}$ satisfies the necessary first-order condition*

$$X\bar{\Lambda}_i = \bar{q}\left(\bar{a}^j\right) + \frac{\mu}{m_f} X \Gamma X' \bar{a}_i.$$

4. *Consumption allocations are invariant to trading a different set of securities that has the same asset span.*

Proposition 1 shows the first-order condition for portfolio optimality of the competitive fringe immediately generates a closed-form demand system for all assets. Specifically, prices are equal to the marginal utility of the fringe evaluated at the equilibrium consumption level. Price impact is then the marginal change in fringe marginal utility induced by a change in strategic agents' quantities. By market-clearing, this can be inferred by simply writing fringe consumption as a function of strategic agents' portfolios. This leads to a closed-form expression for price impact that is linear in the fringe's marginal utility across states because of the quasi-linearity of fringe utility.

Given the pricing system, the optimal portfolio choice of strategic agents equates asset prices to expected state prices across all states in which the asset pays off plus an endogenous distortion from price impact that scales with the position size. This wedge distorts down the demand of buyers to lower prices, and the supply of sellers to raise prices. This leads to unrealized gains from trade in the precise sense that expected state-contingent valuations of marginal changes in consumption are dispersed across buyers and sellers. These wedges appear although there are no exogenous barriers that would prevent the realization of all feasible gains from trade. In the sequel, we examine these wedges more carefully by putting more structure on the set of tradeable assets.

3 Benchmark Without Portfolio Constraints

In this section, we characterize the distortions from market power absent any restrictions on what assets agents can trade. Our analysis yields three key insights. First, rent extraction is most efficient in this special case. This is because it allows strategic agents to implement their optimal distortions state-by-state. Second, the voluntary misalignment of state prices lends itself to a formal equivalence between risk sharing under market power and a fictitious economy in which strategic agents behave competitively but trade a restricted asset span. This allows us to map the risk sharing distortions stemming from market power into an interpretable span of under-traded assets. Third, nonlinear price impact gives rise to an illiquidity externality that will be important for understanding how restricting which assets agents trade through regulation can improve welfare.

3.1 Privately-Optimal Distortions and Externalities

To describe privately optimal distortions without mandates or regulation, we adapt Proposition 1 to the market structure containing the full set of Arrow securities. Because equilibrium allocations are invariant to the particular set of tradeable assets (Proposition 1), the resulting insights apply to any complete market structure.[6]

Proposition 2 (Canonical Arrow-Debreu Economy) *Let X equal the identity matrix (the full set of Arrow securities), and denote the price and quantities of the Arrow security referencing state z by $q(z)$ and $a_i(z)$, respectively. Then prices and price impact are given by $q_j(\vec{a}^j) = q(z) = \Lambda_f(z)$ and $q_j'(\vec{a}^j) = -\Lambda_f'(z) = q'(z) > 0$. The state price for a strategic agent of type i is*

$$\Lambda_i(z) \equiv \frac{\pi(z) u_{i,2}'(y_i(z) + a_i(z))}{u_{i,1}'(c_{i1})},$$

The optimal holdings of Arrow security z satisfies the first-order condition

$$\Lambda_i(z) = q(z) + \frac{\mu}{m_f} q'(z) a_i(z) \qquad \text{for all } z.$$

Sellers of assets (agents with $a_i(z) < 0$) increase their position relative to the competitive economy, while buyers (agents with $a_i(z) > 0$) reduce their position. Hence, marginal valuations are

[6] Carvajal (2014) shows no arbitrage need not hold when strategic agents trade with price impact in financial markets, and a competitive fringe is one means of enforcing no arbitrage. The fringe in our setting lets us generalize our insights from a particular market structure to any with an equivalent asset span.

misaligned between buyers and sellers in every state of the world, and consumption allocations are constrained inefficient.

As is standard in financial markets with imperfect competition, agents implement rent-seeking distortions at the cost of under-diversifying their portfolios. The extent to which an agent distorts her holdings of a given asset depends on her price impact in that market $q'(z)$ and the underlying gains from trade $|\Lambda_i(z) - q(z)|$, which represents the scope for rents. Because markets are complete, these wedges can be chosen state-contingently. A direct implication is that equilibrium allocations are constrained inefficient because marginal valuations are misaligned state by state. In particular, trading one more unit of state-contingent consumption between any pair of buyers and sellers of a given asset would lead to strict welfare gains, but these gains from trade are not realized because individual investors strategically internalize their price impact.

The inefficiency of trading arrangements in this economy is further amplified by an *illiquidity externality* that operates through the interaction of strategic agents' demands on price impact. This effect is most transparent with a second-order approximation to the pricing function from Proposition 2 for the asset referencing state z:

$$\Delta q(z) \approx q'(z) \sum_i \Delta a_i(z) + \frac{1}{2} q''(z) \sum_i \sum_k \Delta a_i(z) \cdot \Delta a_k(z) \tag{5}$$

The first linear term on the right-hand side of (5) is the direct price change resulting from change in agent demands evaluated at the slope of the pricing function. This direct effect is also present in models of strategic trading that feature affine prices, such as those analyzed in the CARA-normal tradition of Kyle (1989). With increasing fringe utility, $q'(z) > 0$, and an increase in demand / reduction of supply by any strategic agent raises the asset price.

The second-order term for (5) represents the illiquidity externality that arises only because our model allows for non-linear price impact. It reflects the strategic interaction in the trades of any pair of strategic agents and arises because the asset price is non-linear in the fringe's consumption. If the fringe has convex marginal utility, $q''(z) > 0$, then buyer demand reductions are strategic substitutes while those by sellers are strategic complements. If a buyer reduces her demand, this lowers price impact for another buyer and induces that buyer to purchase more. In contrast, if a seller reduces her supply, price impact increases, which raises the marginal benefit for another seller to reduce her supply. This externality provides a motive for regulatory restrictions on what assets strategic agents trade.

3.2 Aggregate Consequences of Strategic Distortions

In the previous section, we showed that privately-optimal portfolio distortions because of price impact can be cast as a set of state-contingent wedges that capture the private marginal benefit of rationing trades. These make clear that optimal distortions are sensitive to an agent's state-contingent income, her trading needs, and her price impact. However, because these wedges are defined at the individual level, they provide limited insight into the general equilibrium distortions that they entail, which is the key object of interest for policy.

To measure these aggregate consequences, we now map state-contingent wedges into market-wide risk-sharing arrangements by constructing a map between the equilibrium of our model with complete markets and an *equivalent counterfactual economy with competitive trading but incomplete markets*. Such a mapping is feasible because market power leads to misaligned marginal valuations (i.e., state prices) similar to what occurs in models with incomplete markets. The particular form of incompleteness of the counterfactual asset span (i.e. the markets which appear to be missing) then provides a clean measure of the directions in which trade is distorted by strategic trading.

We use an approach similar to Constantinides and Duffie (1996). Specifically, we construct an equivalence between the consumption allocation of strategic agents determined in the market equilibrium in Proposition 2 and one in a counterfactual economy with competitive trading. We do so in two steps. First, we endow all agents in the counterfactual economy with the consumption allocation they obtain in the market equilibrium (i.e., c_{i1} at date 1 and $c_i(z)$ at date 2). Second, we consider the set of market structures that generate no further trade away from this allocation if agents behave competitively. Although there always exist such market structures, such as autarky, we show the *maximum rank* restricted asset span is less than full rank. This is because any unrealized gains among agents who trade competitively can only be explained by exogenous barriers to trade, i.e., missing assets. Indeed, if there were no unrealized gains from trade, the maximum-rank restricted asset span would be complete markets.

The formal definition of such a counterfactual economy is as follows.

Definition 3 (Counterfactual Economy with Competitive Trading) *A counterfactual economy with competitive trading of rank M consists of a state-contingent endowments process $\mathbf{E} = \{c_i 1, \{c_i(z)\}_{i=1}^N\}$ for all agents at all dates and an $M \times Z$ asset return matrix $\tilde{x}$ such that:*

1. *The endowment process is the consumption process from the equilibrium in Proposition 2.*

2. *All agents solve their decision problem (3) taking the payoff matrix and all prices as given.*

3. *No trade is a solution to all agents' decision problems.*

To construct the restricted asset span, we collect all the no-arbitrage conditions among the N types of strategic agents based on their state prices from the market equilibrium, $\Lambda_i(z)$, but now assume they price assets competitively. Given a market structure, indexed by a matrix $\tilde{x}$ of M asset returns by security and state, this will give rise to a matrix equation based on the N strategic agent type's Euler Equations to solve for each $\tilde{x}_m \in \tilde{x}$. This $\tilde{x}$ is the fictitious restricted asset span. The goal is to find the largest M for which such a solution with linearly independent asset returns exists.[7]

In what follows, let $\mathcal{M}$ be the $N \times Z$ matrix of strategic agents' state prices (i.e., $\mathcal{M}_{iz} = \Lambda_i(z)$) and $\lambda_i(z) = \Lambda_i(z) / \pi(z)$ be agent i's stochastic discount factor (SDF). In addition, let ι_N be the $N \times 1$ vector of ones and superscript T denote the transpose. We then have the following proposition.

Proposition 3 *There exists a counterfactual economy with competitive trading in which the return matrix $\tilde{x}$ has maximal rank M less than the number of states Z (i.e., $\tilde{x}$ is rank deficient), where $\tilde{x}$ is the largest rank matrix whose columns satisfy:*

$$\mathcal{M}\tilde{x}_m = \iota_N \qquad \text{for all } m$$

It is sufficient (although not necessary) that second moment matrix of agents' state prices ($\mathcal{M}\mathcal{M}^T$) has full rank for $\tilde{x}$ to have a nontrivial solution. The extent of implied market incompleteness, as measured by unrealized gains from trade, satisfies:

$$Cov\left(\lambda^*(z), \lambda_i(z) - \lambda_{i'}(z)\right) = 0,$$

where $\lambda^*(z)$ (given in (36)) is the projection of the SDF onto the incomplete return space.

[7]We do not need to include the competitive fringe in this calculation because we can use the fringe's state prices to pin down asset prices once we have recovered the return matrix $\tilde{x}$. This is because $\tilde{x}$ is a dividend yield (dividend x divided by price p), and we are free to specify the two separately such that the fringe's no arbitrage conditions hold.

The second part of Proposition 3 states that the restricted asset span is such that any potential gains from trade are unpriced by the unique state price deflator implied by market returns. When markets are incomplete, one can always recover such a market-implied deflator (e.g., Hansen and Jagannathan (1991)). Any unrealized gains from trade are consequently interpreted as untradeable under the hypothesis of perfect competition. Our analysis inverts the insights of Hansen and Jagannathan (1991) to show one can recover the implied asset span from the cross-section of traders' state prices.

The key implication of this construction is the restricted asset span provides a taxonomy of risks that can be fully shared under market power (i.e., lie within the span) and those not shared enough (i.e., are orthogonal to it). The former primarily facilitate rent-seeking while the latter are under-traded because strategic agents retain diversifiable risk. In the next subsection, we implement this technology using two canonical settings.

3.3 Canonical Settings

We now complement our general analysis by illustrating the basic mechanisms in specific settings where the fundamental sources of gains from trade, and thus efficient trading patterns, are well understood. To do this, we specialize our model to two canonical settings. In the first, ex-ante symmetric agents face purely diversifiable risk. In the second, agents with different risk aversion can trade pure aggregate risk. We illustrate both privately optimal distortions in complete markets and the equivalent incomplete markets economy. We later derive optimal regulatory mandates in these settings as well.

3.3.1 Canonical Setting 1: Pure Risk Sharing

We use the following setting to discuss risk sharing among ex-ante symmetric investors.

Setting 1 (Pure Diversifiable Risk) *All strategic agents have the same concave utility functions u with convex marginal utility, and the fringe's utility function at date 2 is also given by u. There are two types of strategic agents, $i \in \{1,2\}$. All strategic agents and the fringe have an initial endowment of $\bar{y}$. At date 2, there are two possible states $z \in \{1,2\}$ with $\pi(z) = \frac{1}{2}$. Strategic agents face pure idiosyncratic risk: $y_i(i) = \bar{y} + \Delta$ and $y_i(3-i) = \bar{y} - \Delta$, i.e., in every state one type has high and the other has low income. The fringe receives $\bar{y}$ in every state.*

Because strategic agents are ex-ante symmetric, we search for an equilibrium where each strategic agent sells $a_S < 0$ claims on the state in which she has high income, and buys a_B units of the claim on the state in which she has low income. The consumption of the fringe is constant across states, both securities have the same prices $q^* = \frac{1}{2}u_f'(c_{f2}^*)$ and price impact $q^{*\prime} = -\frac{1}{2}u_f''(c_{f2}^*)$, and strategic agents consume the same at date 1.

Privately optimal distortions. Perfect risk sharing between the two types would require $a_S = -\Delta$ and $a_B = \Delta$. We therefore write the optimal security positions as $a_S = -\Delta + \delta_S$ and $a_B = \Delta - \delta_B$, where δ_S and δ_B are optimally chosen deviations from perfect risk sharing when selling and buying, respectively. By the first-order conditions, buyers and sellers choose optimal wedges between marginal valuations and state prices

$$\text{Seller distortion:} \qquad \left| \frac{\frac{1}{2}u'(\bar{y}+\delta_S)}{u'(\bar{y}+q^*\cdot(\delta_S-\delta_B))} - q^* \right| = \frac{\mu}{m_f} q^{*\prime}\Big(\Delta - \delta_S\Big) \tag{6}$$

$$\text{Buyer distortion:} \qquad \left| \frac{\frac{1}{2}u'(\bar{y}-\delta_B)}{u'(\bar{y})-q^*\cdot(\delta_S-\delta_B))} - q^* \right| = \frac{\mu}{m_f} q^{*\prime}\Big(\Delta - \delta_B\Big). \tag{7}$$

Buyer and seller valuations are misaligned state-by-state although agents are symmetric ex ante, where the right-hand side summarizes the lost gains from trade.

Asymmetric distortions and price impact. Although strategic agents are ex-ante symmetric, sellers distort their positions *more* than buyers, i.e., $\delta_S > \delta_B > 0$.[8] This is because a seller starts with higher income than a buyer, and therefore her marginal utility of additional consumption is strictly lower. Because the fringe's marginal utility is convex in its consumption, prices and price impact rise because of market power.

Equivalent restricted asset span. Since strategic agents are ex ante symmetric, there exist two distinct state prices Λ^l and $\Lambda^h > \Lambda^l$ such that $\Lambda_i(i) = \Lambda^l$ and $\Lambda_i(-i) = \Lambda^h$. Every agent assigns a low marginal value of consumption to the state with high private returns, and a high marginal value of consumption to the state with low private returns. Because there are two states, the equivalent restricted asset span has a single asset. The construction from Proposition 3 shows the dividend-yield $\{\tilde{x}_z\}_{z=1,2}$ satisfies

$$\begin{bmatrix} \Lambda^h & \Lambda^l \\ \Lambda^l & \Lambda^h \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

[8] If instead $\delta_S = \delta_B$, then the right-hand side is the same for buyers and sellers. However, the left-hand side is strictly smaller for sellers than buyers for any given distortion; as such, $\delta_S > \delta_B > 0$.

Solving this equation gives

$$\tilde{x}_1 = \tilde{x}_2 = \frac{1}{\Lambda^h + \Lambda^l} = r_f^*.$$

This is a risk-free bond whose return is the inverse sum of state prices (the risk-free rate).

With two types and diversifiable risk, consumption allocations are such that an outside observer would infer that only a risk-free bond can be traded. This is because a risk-free bond only allows agents to shift resources across time, whereas there are gains from trade in shifting resources across states. Because these risks are not fully shared, the outside observer interprets this as prima facie evidence this risk cannot be traded. The "rationed asset" induced by market power is then the asset whose return is orthogonal to a risk-free bond that, if it were available under perfect competition, would allow agents to realize the residual gains from trade. In this example, it is a simple swap.

Corollary 1 (Missing Asset with Pure Idiosyncratic Risk) *A security with payoff* $[1\ x]$ *is orthogonal to the risk-free bond if and only if*

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & x \end{bmatrix}' = 0.$$

This requires $x = -1$*. The "missing asset" that would have allowed for perfect insurance is therefore an idiosyncratic risk swap.*

3.3.2 Canonical Setting 2: Heterogeneous Preferences

We now introduce the basic setting in which we study heterogeneous preferences.

Setting 2 (Heterogeneous Preferences) *There are two types of strategic agents,* $i \in \{1, 2\}$*, and two states of the world,* $z \in \{l, h\}$*. There is only aggregate risk,* $y_i(h) = y_h$ *and* $y_i(l) = y_l$ *for all* i*. Risk attitudes are heterogeneous: Type 1 is strictly risk-averse and Type 2 is risk-neutral. The fringe has the same utility over date 2 consumption as the risk-averse agent.*

Given that one strategic type (and competitive fringe) is risk averse, it is efficient for the risk-neutral agent to hold all risk exposure. Under market power, however, risk sharing is imperfect because the risk neutral agent restricts her supply of claims on the low aggregate state. As such, the risk-neutral agent has a constant state price for all states, $\Lambda_2(z) = \Lambda_{rn} > 0$, while the risk-averse agent has a higher state price in the low state. In particular, this agent has two distinct state prices satisfying $\Lambda^h < \Lambda_{rn}$ and $\Lambda^l > \Lambda_{rn}$.

Equivalent restricted asset span. With two aggregate states, the implied restricted asset span from Proposition 3 has one asset with dividend-yield $\{\tilde{x}_z\}_{z=l,h}$ satisfying

$$\begin{bmatrix} \Lambda_{rn} & \Lambda_{rn} \\ \Lambda^h & \Lambda^l \end{bmatrix} \begin{bmatrix} \tilde{x}_h \\ \tilde{x}_l \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Solving this equation gives

$$\tilde{x}_h = \frac{\Lambda^l - \Lambda_{rn}}{\Lambda_{rn}\left(\Lambda^l - \Lambda^h\right)} \qquad \text{and} \qquad \tilde{x}_l = \frac{\Lambda_{rn} - \Lambda^h}{\Lambda_{rn}\left(\Lambda^l - \Lambda^h\right)}.$$

The asset carries exposure to aggregate risk since $\tilde{x}_h > \tilde{x}_l$ by Jensen's inequality. Therefore, it is a levered market index.

Because the risk-neutral agent rations insurance against aggregate shocks, the outside econometrician would conclude there are no assets that offer sufficient protection against aggregate risk. This has a natural interpretation: the risk-averse agent can only trade a market index with a degree of risk exposure that leaves no further insurance opportunities. Similar to the case of pure diversifiable risk, we can find the "rationed" asset by searching for the asset whose payoff is orthogonal to the levered market index. This missing asset is similar to a put on the stock market.

Corollary 2 (Missing Asset with Pure Aggregate Risk) *A security with payoff $[1\ x]$ is orthogonal to the levered equity portfolio if and only if*

$$\left[\begin{array}{cc} 1 & \frac{\Lambda_{rn}-\Lambda^h}{\Lambda^l-\Lambda_{rn}} \end{array}\right] \left[\begin{array}{cc} 1 & x \end{array}\right]' = 0.$$

This requires $x = -\frac{\Lambda^l-\Lambda_{rn}}{\Lambda_{rn}-\Lambda^h} < 0$. We can rescale this vector as $\left[\begin{array}{cc} \Lambda_{rn} - \Lambda^h, & \Lambda_{rn} - \Lambda^l \end{array}\right]$, which is an aggregate risk swap that pays off in the high state and loses money in the low state.

In the next section, we explore how imposing constraints on what they can trade can reallocate or even mute their ability to exert market power.

4 Effects of Regulatory Mandates

In Section 3, we established how large institutional investors strategically ration their trading of assets valuable for risk sharing to extract inframarginal rents. In this section, we investigate how imposing mandates, or constraints on what these agents can trade, alters their ability to exert market power. To preserve the tractability of our basic framework, we maintain the assumption that the underlying asset markets are composed of Arrow securities. We can then model mandates as constraints on the combinations of Arrow securities that a particular agent can trade. For example, an agent may be forced to take some position $a_i(z) = \psi(z)a_i(z^*)$ in Arrow security z if she takes a position $a_i(z^*)$ in Arrow security z^*, where $\psi(z)$ represents the constraint. This leads to a flexible specification that can be used to study the design of mandates.

In practice, mandates are often imposed privately at the institution-level through investment mandates,[9] and publicly at the industry-level through regulation. Our main positive insight is that once general equilibrium forces are taken into account, investment mandates and regulation limited to only a subset of market participants reallocates market power toward unconstrained strategic agents, which can have an ambiguous impact on liquidity and welfare (Section 4.1). In contrast, regulation that jointly restricts what all strategic agents can trade can improve both liquidity and welfare (Section 4.2).

Our main normative result is Proposition 4, which establishes that there generically exists a restricted asset span with *better* risk sharing than without mandates or regulation. By bundling states together in a limited set of securities, regulation that restricts what institutional investors can trade blunts their incentives to sub-optimally manage risks state-by-state to extract rents. We then use our two canonical settings with either ex-ante symmetric agents or heterogeneous preferences to derive insights into the optimal design of market-wide regulatory mandates. We show that these are directly linked to the set of *rationed assets* constructed in the equivalent incomplete markets economy from Section 3. We leave the optimal design problem for general preferences and payoffs to future work.

4.1 Mandates on Specific Investors

We begin our analysis of mandates by analyzing the positive effects of restricting agents to trade only certain combinations of Arrow securities. Since these effects are most transparent in settings where the gains from trade are clear and well-understood, we focus

[9]For instance, such mandates may be put in place because of moral hazard concerns (e.g., He and Xiong (2013)). Gabaix and Koijen (2021) examine the asset pricing implications of investment mandates among competitive financial institutions.

on the canonical settings introduced in Section 3.3 and study mandates closely related to the missing asset span discussed in Section 3.2. In the first, we show how mandates can mute strategic considerations. In the second, we show that they may redistribute market power.

4.1.1 Pure Risk Sharing

We begin with Setting 1, in which ex-ante symmetric agents share diversifiable risk. We impose portfolio constraints on a subset of large agents by assuming a fraction χ of each type is exogenously restricted to take a position in asset 2 that is the negative of her position in asset 1, i.e. $a(2) = -a(1)$. Observe this is equivalent to forcing these agents to trade only an asset with payoffs $[1, -1]$, but allows for direct comparisons without mandates for unaffected agents.

For restricted agents, perfect risk sharing at date 2 requires $a(1) = \Delta$ if the agent is of type 2 and $a(1) = -\Delta$ is of type 2. Let δ^{TR} denote the absolute deviation from perfect risk sharing for an agent with trading restrictions (that is, $a(1) = \Delta - \delta^{TR}$ if the agent is type 2 and $a(1) = -\Delta + \delta^{TR}$ if the agent is of type one). Because trading restrictions enforce zero net expenditures on financial claims at date 1, their first-order condition is

$$\frac{1}{2}\left|\frac{\frac{1}{2}u'(\bar{y}+\delta^{TR})}{u'(\bar{y})} - q^*\right| + \frac{1}{2}\left|\frac{\frac{1}{2}u'(\bar{y}-\delta^{TR})}{u'(\bar{y})} - q^*\right| = \frac{\mu}{m_f} q^{*\prime}\left(\Delta - \delta^{TR}\right),$$

We now show that these restrictions induce weaker incentives to distort, and that this can mute the illiquidity externality.

Weaker incentives to distort. To understand whether restrictions weaken strategic considerations, we compare optimal policies of restricted agents to those of investors not subject to mandates. Holding prices and price impact fixed, the denominator of state prices is strictly higher when agents face trading restrictions. This is because the trading restrictions enforce zero net expenditures on assets at date 1, leading them to consume $\bar{y}$ rather than $\bar{y} - q^*(\delta_S - \delta_B)$. This raises the cost of distorting at date 2, and therefore leads to lower average distortions, $\delta^{TR} < \frac{1}{2}(\delta_S + \delta_B)$.

Muting the illiquidity externality. Trading restrictions also reduce price impact by neutralizing the illiquidity externality. Agents with trading restrictions have zero net demand for consumption at date 2 (that is, they share risk among themselves, but not with the fringe). Fringe consumption is therefore $c_{f2} = \bar{y} - \frac{1-\chi}{m_f}(\delta_S - \delta_B)$, which is strictly increasing in χ. Trading restrictions therefore further attenuate the distortions from market power by diluting price impact. The limitation of this result, however, is that only a subset of agent face constraints in this example. In the next subsection, we consider direct changes to the asset span traded by all agents.

4.1.2 Heterogeneous Preferences

We now consider the case with heterogeneous preferences and pure aggregate risk case from Canonical Setting 2, and study the effects of asymmetric mandates in which the risk-neutral type is constrained but the risk-averse is not. We show that constraining only a subset of strategic agents can reallocate market power to those that are unconstrained.

Given the environment from Setting 2, suppose that the risk-neutral type 2 agent is restricted to take a position in one asset that is the negative of her position in the other , i.e., $a_2(l) = -a_2(h) = a_2$. We further assume both states are equally likely, that $u'(1) = 1$ for the risk-averse type 1 agent's preferences (i.e., log or CRRA), and initial wealth is $\bar{y} = \frac{y_h + y_l}{2}$. With perfect risk sharing, $q(h) = \frac{1}{2}$, $q(l) = \frac{1}{2}$, and type 1 agents can buy $a_1(h) = \frac{1}{2}(y_l - y_h) < 0$ state h claims and $a_1(l) = -a_2(h) > 0$ state 2 claims. Suppose further the fringe has endowment 1 in each state; as a result, the risk-neutral agent takes the off-setting positions $a_2 = \frac{1}{2}(y_l - y_h)$.

Let δ^2 denote the absolute deviation from perfect risk sharing for the type 2 agent with trading restrictions, i.e., $a_2 = \frac{1}{2}(y_l - y_h) + \delta^2$. The first-order condition for the risk-neutral type 2 agent is

$$q^*(h) - q^*(l) = \frac{\mu}{m_f}\left(q^{*'}(l) + q^{*'}(h)\right)\frac{1}{2}(y_l - y_h) + \frac{\mu}{m_f}\left(q^{*'}(l) + q^{*'}(h)\right)\delta^2.$$

The risk-neutral agent effectively trades a zero payoff swap with price $q^*(l) - q^*(h)$ and price impact $\frac{\mu}{m_f}\left(q^{*'}(h) + q^{*'}(l)\right)$. Substituting this condition into the first-order condition for the risk-averse agent for the low minus the high state claim gives

$$\frac{\frac{1}{2}u'\left(\bar{y} - \delta_l^1\right) - \frac{1}{2}u'\left(\bar{y} + \delta_h^1\right)}{u'\left(\bar{y} + \frac{(q^*(l) - q^*(h))(y_h - y_l)}{2} + q^*(l)\delta_l^1 - q^*(h)\delta_h^1\right)} = \frac{\mu}{m_f}\left(q^{*'}(l)\left(\delta_l^1 - \delta^2\right) + q^{*'}(h)\left(\delta_h^1 - \delta^2\right)\right).$$

Mixed incentives to distort. Because the risk-neutral type must take off-setting positions in the two Arrow assets, her incentives to distort are blunted because she cannot simultaneously reduce her demand and supply in both markets (i.e., δ^2 is the same in both markets). In contrast, the risk-averse type has stronger incentives to distort. First, she can reduce her supply more for the high state Arrow asset than her demand for the low Arrow asset (i.e., $\delta_h^1 > \delta_l^1$). Second, the risk-neutral type cannot raise the price of the low state $q^*(l)$ and lower the price of the high state Arrow asset $q^*(h)$ as much as without the portfolio restrictions, which reinforces the risk-averse type's market power.

Asymmetric impact on the illiquidity externality. Because the risk-neutral agent exerts less market power with the portfolio restriction, she cannot raise the price of the low state claim $q^*(l)$ as successfully. As a result, $q^*(l)$ is lower and because the fringe has convex marginal utility, price impact is also lower. In contrast, the risk-averse agent can now more successfully raise the price of the high state claim $q^*(h)$ because the risk-neutral type cannot reduce her demand as much. As a result, $q^*(h)$ is higher, and consequently so is price impact for that market. Restricting only one agent type can consequently improve liquidity in one market while worsening it in another. In practice, this means that mandates on regulated institutions such as pension funds may reallocate pricing power to investors with fewer constraints, such as hedge funds.

Overall, these two settings illustrate the impact of investment mandates and narrow portfolio regulation in oligopolistic financial markets. The first example illustrates that such restrictions when applied to a subset of all strategic agent types reduces incentives to distort and can improve market liquidity. In contrast, the second example demonstrates that when only one agent type is constrained, that agent forfeits her exertion of market power to other agent types. The impact on market liquidity in this scenario is mixed: it improves in markets in which she is a seller but can worsens in markets in which she is a buyer. This further has distributional consequences for welfare.

4.2 Market-wide Mandates

In this subsection, we examine whether jointly restricting which assets all large institutional investors can trade through broad regulation improves welfare by facilitating better risk sharing. As described above, the main motive for such regulation is to improve risk sharing and to mute the illiquidity externality.

Let $U_j(X, \mu)$ be the expected utility obtained by agent i given asset span X and market concentration μ. Our welfare criterion is size-weighted utilitarian welfare,

$$\mathcal{W}(X,\mu) = \sum_i U_i(X,\mu) + m_f U_f(X,\mu)$$

The next proposition is our key result: it is generically possible to increase welfare by restricting the asset span large agents can trade as long as there are strictly fewer types of strategic agents than states of the world.

Proposition 4 *If $N < Z$, there generically exists a restricted asset span that delivers higher social welfare than complete markets.*

Market power leads to first-order welfare distortions because large agents' state prices are misaligned state-by-state. With no restrictions on portfolios, we can always replicate the equilibrium allocation with a set of bespoke assets tailored to each agent. When there are fewer types than states, this is trivially a restricted asset span. What is nontrivial is because marginal valuations are misaligned in complete markets, it is possible to engineer a welfare improvement by twisting the payoffs of these bespoke assets. As a result, welfare with an optimally chosen restricted asset span must be *strictly* higher than without portfolio restrictions. By reallocating consumption from an agent with low to one with high marginal utility in a particular state, changing the asset span leads to an improvement in social welfare.

The key insight is that letting large institutional investors trade all risks in complete markets maximizes the scope for rent extraction by permitting those with market power to choose optimal state-contingent distortions. Although restricting what assets these agents can trade has the direct cost of making it more difficult to trade, it has the indirect effect of muting market power. This result shows the indirect benefit exceeds the direct cost for suitably chosen forms of restricted asset spans.

4.3 Optimal Mandates

In this section, we characterize welfare-improving forms of portfolio regulation in our two canonical settings with diversifiable and aggregate risk. Specifically, we show a swap, which is the "rationed asset" from our analysis in Section 3, improves risk sharing and welfare relative to no mandates or regulation in both settings. To focus on the asymmetric incentives of strategic agents to distort in the aggregate risk example, we focus on the limit in which both strategic agents and the competitive fringe become arbitrarily small (i.e., $\mu, m_f \to 0$), but price impact survives in the limit (i.e,. $\frac{\mu}{m_f} \to \kappa$ a constant). Neuhann and Sockin (2023) provides a formal analysis of this particular limit economy.

Proposition 5 *The optimal asset span in our canonical examples can be described as follows:*

(i) Suppose $u'(\bar{y}) \geq 1$. With pure idiosyncratic risk, welfare is strictly higher in restricted asset span where the only asset is a swap that pays $[1, -1]$ than in complete markets.

(ii) Consider the limit of the economy as $m_f \to 0$, holding μ/m_f fixed. With pure aggregate risk in which one type of strategic agent is strictly less risk averse than the other and the date-2 utility function of the competitive fringe is the same as that of the more risk-averse strategic type, there exists an aggregate risk swap with payoffs $[x_L, x_H]$ and $x_H x_L < 0$ for which social welfare is higher than with complete markets.

The intuition for the diversifiable risk case is straightforward. With ex-ante symmetric agents, trading a swap is sufficient to realize all gains from trade. But because agents are ex-post heterogeneous, a seller has stronger incentives to distort than a buyer. This is because her marginal utility schedule is flatter. By bundling states, the swap lowers the incentives to distort as a seller, which improves risk sharing and liquidity. The aggregate risk case follows similar logic. Although the risk-neutral agent can exert more market power in all states, she does so particularly where the risk-averse agent is desperate for additional consumption (i.e., the low state). By fixing the terms of trade across high and low states, the restricted asset span leads to weaker distortions.

A general insight from our analysis is that investment mandates and narrow portfolio regulation imposed on only a subset of large institutions may have the unintended effect of reallocating those institutions' market power to their trading partners with distributional consequences for liquidity and welfare. In contrast, broad portfolio regulation can improve risk sharing and limit the rent-seeking behavior of large financial institutions when it restricts the assets they can trade only to those that are essential for hedging their business risks. Incrementally allowing these investors to trade other illiquid assets improves their ability to exert market power, and induces them to under-diversify their portfolios to extract surplus from financial markets. Although individually beneficial, socially this is costly because the exertion of market power introduces pecuniary externalities that do not cancel even when large agents are unrestricted in what they can trade. Our insights suggest that ongoing regulatory efforts to relax portfolio restrictions on NFBIs, such as the recent New York and United Kingdom reforms to make it easier for pension funds to invest in alternative assets, may have the unintended consequence of worsening their risk management.

5 Conclusion

In this paper, we explore how restrictions on what large non-bank financial institutions can trade impacts their risk management. Because such investors trade in relatively illiquid asset classes, we analyze this issue through the lens of imperfect competition in financial markets. Our analysis delivers two key insights. First, we show that the portfolio distortions from market power are highly state-contingent, and that risk sharing is most impaired when gains from trade are large and strategic investors are unrestricted in what they can trade. There is, in fact, an equivalence between risk sharing under market power with no restrictions and under perfect competition with portfolio constraints on all investors. As a result, assets can be dichotomized into the types of risks that are shared effectively in financial markets and those that are rationed because of market power. Second, restricting the trading opportunities of large investors through broad portfolio regulation can counter-intuitively improve risk sharing. This is because when all large investors face the same restrictions, portfolio constraints attenuate their rent-seeking behavior. In contrast, privately-imposed investment mandates and narrow portfolio regulation that targets only a subset of non-bank financial institutions may worsen their risk management by reallocating market power to their trading partners. Our analysis suggests the current regulatory trend of relaxing portfolio constraints on these investors may be counterproductive in improving their risk management.

References

ALVAREZ, F., AND U. JERMANN (2000): "Efficiency, Equilibrium, and Asset Pricing with Risk of Default," *Econometrica*, 68(4), 775–797.

ATHANASOULIS, S., AND R. SHILLER (2000): "The Significance of the Market Portfolio," *Review of Financial Studies*, 13, 301–329.

BABUS, A., AND K. HACHEM (2020): "Markets for Financial Innovation," *NBER Working Paper 25477*.

——— (2021): "Regulation and Security Design in Concentrated Markets," *Journal of Monetary Economics*, 121(1), 139–151.

BABUS, A., AND C. PARLATORE (2021): "Strategic Fragmented Markets," *Journal of Finance Economics*.

BASAK, S. (1997): "Consumption Choice and Asset Pricing with a Non-price-taking Agent," *Economic Theory*, 10(3), 437–462.

BASAK, S., AND A. PAVLOVA (2013): "Asset Prices and Institutional Investors," *American Economic Review*, pp. 1728–58.

BLUME, L. E., T. COGLEY, D. A. EASLEY, T. J. SARGENT, AND V. TSYRENNIKOV (2018): "A Case for Incomplete Markets," *Working Paper*.

BOCOLA, L., AND G. LORENZONI (2020): "Risk Sharing Externalities," *mimeo Northwestern*.

BUCHAK, G., G. MATVOS, T. PISKORSKI, AND A. SERU (2018): "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks," *Journal of Financial Economics*, 130, 453–483.

——— (2022): "Beyond the Balance Sheet Model of Banking: Implications for Bank Regulation and Monetary Policy," *Working Paper*.

CARVAJAL, A. (2014): "Arbitrage Pricing in Non-competitive Financial Markets," *mimeo University of Western Ontario*, (1).

CARVAJAL, A., M. ROSTEK, AND M. WERETKA (2012): "Competition in Financial Innovation," *Econometrica*, 80, 1895–1936.

CASS, D., AND A. CITANNA (1998): "Pareto Improving Financial Innovation in Incomplete Markets," *Economic Theory*, 11, 467–494.

CONSTANTINIDES, G. M., AND D. DUFFIE (1996): "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104(2), 219–240.

DÁVILA, E., AND A. KORINEK (2018): "Pecuniary Externalities in Economies with Financial Frictions," *Review of Economic Studies*, 85.

DEMANGE, G., AND G. LAROQUE (1995): "Optimality of Incomplete Markets," *Journal of Economic Theory*, 65, 218–232.

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

EISENBACH, T., AND G. PHELAN (2022): "Cournot Fire Sales," *American Economic Journal: Macroeconomics*, 14, 508–42.

ELUL, R. (1995): "Welfare Effects of Financial Innovation in Incomplete Markets Economies with Several Consumption Goods," *Journal of Economic Theory*, (65), 43–78.

GABAIX, X., AND R. KOIJEN (2021): "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," *Working Paper*.

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

GREENWALD, B. C., AND J. STIGLITZ (1986): "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics*, 101, 229–264.

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*.

HANSEN, L. P., AND R. JAGANNATHAN (1991): "Implications for Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

HART, O. (1975): "On the Optimality of Equilibrium When the Market Structure is Incomplete," *Journal of Economic Theory*.

HE, Z., AND W. XIONG (2013): "Delegated Asset Management, Investment Mandates, and Capital Immobility," *Journal of Financial Economics*, 107, 239–258.

HELLWIG, C., AND G. LORENZONI (2009): "Bubbles and Self-Enforcing Debt," *Econometrica*, 77(4), 1137–1164.

JANSEN, K. A. (2021): "Long-term Investors, Demand Shifts, and Yields," *Working Paper*.

JANSEN, K. A., S. KLINGER, A. RANALDO, AND P. DUIJM (2023): "Pension Liquidity Risk," *Working Paper*.

KACPERCZYK, M., J. NOSAL, AND S. SUNDARESAN (forthcoming): "Market Power and Price Informativeness," *Review of Economic Studies*.

KEHOE, T. J., AND D. K. LEVINE (1993): "Debt-Constrained Asset Markets," *The Review of Economic Studies*, 60(4), 865–888.

KHETAN, U., I. NEAMTU, AND I. SEN (2023): "The Market for Sharing Interest Rate Risk: Quantities Behind Prices," *Bank of England Staff Working Paper*.

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

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

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

MARIN, J. M., AND R. RAHI (2000): "Information Revelation and Market Incompleteness," *Review of Economic Studies*, 67, 455–481.

NEUHANN, D., AND M. SOCKIN (2023): "Financial Market Concentration and Misallocation," *mimeo UT Austin*.

OECD (2015): "Regulation of Insurance Company and Pension Fund Investment," *OECD Report to G20 Finance Ministers and Central Bank Governors*.

——— (2021): "Annual Survey of Investment Regulation of Pension Funds and Other Pension Providers," .

PESENDORFER, W. (1995): "Financial Innovation in a General Equilibrium Model," *Journal of Economic Theory*, 65, 79–116.

PINTER, G., AND D. WALKER (2023): "Hedging, Market Concentration and Monetary Policy: a Joint Analysis of Gilt and Derivatives Exposures," *Bank of England Staff Working Paper*.

RAHI, R., AND J.-P. ZIGRAND (2009): "Strategic Financial Innovation in Segmented Markets," *Review of Financial Studies*, 22, 2941–2971.

ROSTEK, M., AND J. H. YOON (2021): "Exchange Design and Efficiency," *Econometrica*, 89, 2887–2928.

A Appendix

A.1 Proof of Proposition 1

Step 1: The Problem of the Fringe:

From the first-order condition for the optimal holdings of asset j, a_f^j, from the competitive fringe's problem (4), we can recover the pricing equation:

$$\tilde{q}_j = \sum_{z \in \mathcal{Z}} x_j(z)\, \pi(z)\, u'_{f,2}\left(c_f(z)\right) = \sum_{z \in \mathcal{Z}} x_j(z)\, \Lambda_f(z)\,, \tag{8}$$

where $\Lambda_f(z) = \pi(z)\, u'_{f,2}\left(c_f(z)\right)$ is the state price of the competitive fringe in state z. Since fringe consumption in state z satisfies (4) and a_f^j satisfies market-clearing (2), it is immediate from (8) that price impact satisfies the matrix:

$$\hat{Q}_i\left(\bar{a}^j\right) = \frac{\mu}{m_f} X \Gamma X' \text{ where } \Gamma_{j,z} = -\Lambda'_f(z)\, \mathbf{1}_{\{j=z\}} \geq 0, \tag{9}$$

which is the same for all strategic agents (i.e., price impact is anonymous). We define the equilibrium price function to be $q_j\left(\bar{a}^j\right) = Q_j\left(\bar{a}^j\right)$.

Substituting the market-clearing condition (2) into the fringe's consumption at date 2 from (4), $c_f(z)$ can be expressed:

$$c_{f2}(z) = y_f(z) + \sum_j x_j(z) \left(-\frac{1}{m_f} \sum_{i=1}^{N} a_i^j\right).$$

Step 2: 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:

$$q_j\left(\bar{a}^j\right) = \sum_{z \in \mathcal{Z}} x_j(z)\, \Lambda_f(z) = \sum_{z \in \mathcal{Z}} x_k(z)\, \Lambda_f(z) = q_k\left(\bar{a}^k\right).$$

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

Step 3: The Problem of Strategic Agents:

To economize on notation, we consider a representative agent of type i. Let φ_i be the Lagrange multiplier on strategic agent i's budget constraint. The first-order condition for optimal initial consumption c_{i1} from strategic agent i's problem (3) is:

$$u'_{i,1}\left(c_{i1}\right) - \varphi_i = 0. \tag{10}$$

The first-order condition for optimal asset holdings of asset j, a_i^j, is:

$$\sum_{z\in\mathcal{Z}} \pi\left(z\right) x\left(z\right) u'_{i,2}\left(c_{i2}\left(z\right)\right) - \varphi_i\left(Q_j\left(\bar{a}^j\right) + \hat{Q}_{i,j}\left(\bar{a}^j\right)\bar{a}_i\right) = 0, \tag{11}$$

where $\hat{Q}_{i,j}\left(\bar{a}^j\right)$ is the j^{th} row of $\hat{Q}_i\left(\bar{a}^j\right)$ and $\bar{a}_i$ is the vector of agent i's asset demands. We can rewrite (11) with strategic agent i's state price $\Lambda_i\left(z\right)$, substituting for φ_i with (10):

$$\sum_{z\in\mathcal{Z}} \Lambda_i\left(z\right) x\left(z\right) - Q_j\left(\bar{a}^j\right) - \hat{Q}_{i,j}\left(\bar{a}^j\right)\bar{a}_i = 0. \tag{12}$$

We next recognize that strategic agent i has rational expectations, and her perceived price impact must be her actual price impact. Let $\bar{\Lambda}_i$ is the vector agent i's state prices. Since this price impact function is pinned down by the fringe from (9), and the equilibrium price function is $q_j\left(\bar{a}^j\right)$, (12) becomes:

$$\bar{q}\left(\bar{a}^j\right) = X\bar{\Lambda}_i - \frac{\mu}{m_f}X\Gamma X'\bar{a}_i, \tag{13}$$

because a strategic agent of type i has mass μ.

Since the budget constraint of strategic agent i will hold with equality in equilibrium by efficiency, the optimal asset holdings a_i^j determine both date 1 and date 2 consumption, c_{i1} and $c_{i2}\left(z\right)$, respectively.

We now recognize that no strategic agent would ever take an infinite position in any asset. To see, this notice that optimal asset holdings satisfy the FONC (13), which we rewrite as:

$$X\bar{\Lambda}_i = \bar{q}\left(\bar{a}^j\right) + \frac{\mu}{m_f}X\Gamma X'\bar{a}_i, \tag{14}$$

If a strategic agent takes an infinite negative position in asset j, $a_i^j \to -\infty$, then there are two cases to consider for the right-hand side of (14). First, if another strategic agent takes an off-setting position in the asset, then q_j and q'_j remain positive and well-defined, and the right-hand side tends to $-\infty$ because of the a_i^j term. Second, if the fringe is forced to absorb the supply, then q_j 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_{i1} becomes infinite, $c_{i1} \to \infty$. 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 (14) tends to positive ∞ (the demand is either offset by a strategic agent or absorbed by the fringe through infinitely negative date 2 consumption). 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.

As there is no infinite supply or demand of any security, the consumption of all agents is also bounded. Consequently, we can bound all controls of strategic agent i's problem, $\{c_{i1}, \{a_i^j\}_{j=1}^J\}$, in a closed and bounded set. By the Heine-Borel Theorem, this set is compact.

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 i's choice correspondence set is also continuous in the optimization problem's primitives (i.e., endowment processes and initial wealth). Consequently, the choice correspondence of strategic agent i's problem is continuous and compact-valued.

It then follows because the objective function of strategic agent 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 i exists. As the choice of i was arbitrary, this holds for all agents of type i and all types $i \in \{1, ..., N\}$.

Step 4: 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 asset positions for all strategic agents to an optimal set of asset positions using the market-clearing conditions (2) 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 market power, 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 5: Market Structure Invariance:

Consider a different $|\mathcal{Z}| \times K$ payoff matrix B with an equivalent asset span and asset price vector $q^B\left(\bar{a}^{B,j}\right)$. By this we mean that there exist vectors $\bar{a}_i^B$ such that:

$$X'\bar{a}_i = B'\bar{a}_i^B. \tag{15}$$

Similar exercises can be done the asset positions $\bar{a}_f^B$ of the fringe.

It is immediate from Step 1 that the prices $q^B\left(\bar{a}^{B,j}\right)$ satisfy:

$$q^B\left(\bar{a}^{B,j}\right) = B\bar{\Lambda}_f, \tag{16}$$

and that

$$\hat{Q}^B\left(\bar{a}^{B,j}\right) = B\Gamma^B B' \quad \Gamma^B_{j,k} = -\Lambda'_f(z)\,\mathbf{1}_{\{k=l\}}.$$

Consider now the analogue of the first-order necessary conditions for optimal asset demand of strategic agent i from Step 2:

$$q^B\left(\bar{a}^{B,j}\right) = B\bar{\Lambda}_i + B\Gamma^B B'\bar{a}_i^B. \tag{17}$$

Suppose now that the consumption allocations of strategic agents and the fringe are the same under asset span B as under X. This implies that $\Gamma^B = \Gamma$. Notice then we can manipulate equation (17), substituting with (16) and (15):

$$\bar{\Lambda}_f = \bar{\Lambda}_i + \Gamma B'\bar{a}_i^B = \bar{\Lambda}_i + \Gamma X'\bar{a}_i. \tag{18}$$

Multiplying by X, we arrive at:

$$q\left(\bar{a}^j\right) = X\bar{\Lambda}_i + X\Gamma X'\bar{a}_i,$$

which is the first-order condition for optimal asset demands under asset span X. Consequently, strategic agents choose the same optimal state-contingent exposures as under both asset spans.

Finally, it remains to be shown that the portfolio of strategic agent i costs the same under asset span X as under B. This is trivial, however, because:

$$q\left(\bar{a}^{B,j}\right)'\bar{a}_i^B = \bar{\Lambda}_f' B' \bar{a}_i^B = \bar{\Lambda}_f' X' \bar{a}_i = q\left(\bar{a}^j\right)'\bar{a}_i,$$

as required. Consequently, equilibrium allocations are invariant to the market structures that implement the same asset span.

A.2 Proof of Proposition 3

Step 1: The Fictitious Asset Span:

We begin with the first part of the claim. Consider the Cournot-Walras Equilibrium allocation of strategic agents of type i in the economy with imperfect competition, $\left\{\left(c_{i1}, \{c_i(z)\}_{z\in\mathcal{Z}}\right)\right\}_{i=1}^N$. Define:

$$\Lambda_i(z) = q(z) + \frac{\partial q(z)}{\partial a_i(z)} a_i(z) \ \forall \ (i,z), \tag{19}$$

to be the implied state price deflator of agent i in state z.

Now consider a fictitious incomplete-markets economy in which all agents behave competitively and take prices as given. In this fictitious incomplete-markets economy, all agents have the same state prices that they have in the market equilibrium, except now we counterfactually assume that they traded competitively.

Our goal is to find an equivalent implied market structure, indexed by a set of $M \le Z$ securities with a $M \times Z$ return matrix $\tilde{x}$ that spans the M (linear combinations of the Z) states, that justifies their ex post dispersion in state prices if these were the assets the agents counterfactually traded. Since our model is static, a security's return is just its dividend yield (i.e., $\vec{\tilde{x}}_m = \vec{x}_m / p_m$ for dividend process $\vec{x}_m$ and price p_m of security m). We will derive the dividends and prices separately later with the competitive fringe. Because the synthetic assets can be derivatives, $\tilde{x}$ can also have negative entries.

The no arbitrage condition for a competitive agent i in security m is the standard Euler Equation:

$$\sum_{z\in\mathcal{Z}} \Lambda_i(z)\,\tilde{x}_m(z) = 1, \tag{20}$$

which we stack into the matrix equation:

$$\mathcal{M}\vec{\tilde{x}}_m = \iota_N, \tag{21}$$

where ι_N is the vector of ones of length N. For a given security m, there is one condition for each of the N types of strategic agents, giving rise to N conditions for each security and $N \times M$ conditions total. As we will show later, we do not need to include the fringe in this construction. This is because their no arbitrage conditions will be trivially satisfied by the definition of the fictitious incomplete-markets equilibrium because we have the freedom to specify security prices p after recovering $\tilde{x}$.

Our goal is to find as many linearly independent solutions to Equation (20) as is feasible, each corresponding to a different security. The maximum number of securities we can recover is our M, i.e., $M = \max_{k \in \{0,\ldots,Z\}} \{rank\left(\tilde{x}\right) = k\}$. We want the maximum number of securities because, if there were an additional asset that replicated the asset span, introducing it would not initiate trade because it would already be priced at its no arbitrage value by all strategic agents. Notice that the rank of $\mathcal{M}$ satisfies $rank\left(\mathcal{M}\right) \leq min\{N, Z\}$. If $N < Z$, then the system is always under-identified and can have many solutions, while if $N = Z$, then it may have a unique solution if it is identified. Since we assume $N \leq Z$, these are the two cases that we consider in the paper. If instead $N > Z$, it may be over-identified and have no solution (in which case $\tilde{x}$ is the empty set).

To see the content of Equation (21), suppose that all agents were actually competitive instead of strategic. In this case, their state prices would be aligned ex post (i.e., $\Lambda_i\left(z\right) = \Lambda_{i'}\left(z\right) = \Lambda\left(z\right)$ for all i, i'), and $\mathcal{M}$ would reduce to $\iota_N \vec{\Lambda}^T$, where $\vec{\Lambda}$ is the vector of unique state prices. In this case, we can stack Equation (21) across M securities to find:

$$\iota_N \vec{\Lambda}^T \tilde{x}^T = \iota_N \iota_M^T. \tag{22}$$

and Equation (22) reduces to:

$$\tilde{x}\vec{\Lambda} = \iota_M. \tag{23}$$

It is clear then that a solution to Equation (23) is $\tilde{x}^T = diag\left(\vec{\Lambda}\right)^1$, where $diag$ is the diagonal operator on the vector $\vec{\Lambda}$. As is immediate, this corresponds to the complete market with Arrow-Debreu securities each with payoff 1 and price $\Lambda\left(z\right)$ for state z. Notice that this argument does not rely on the number of agents, so state price dispersion from market power is needed for the fictitious asset span not to be complete markets.

With this return matrix $\tilde{x}$, we have constructed a restricted asset span that measures the degree of market incompleteness by replicating the effective asset span of the Cournot-Walras Equilibrium with complete markets. Since at least one state price is misaligned among large agents, $rank(\tilde{x}) < Z$, and markets must be incomplete.

We next establish a sufficient condition that ensures a nontrivial solution to Equation (23). By the Fredholm Alternative, either Equation (21) has a solution or there exists a nontrivial y such that:

$$\mathcal{M}\vec{y} = \vec{0}_Z \qquad \text{and} \qquad \vec{y}^T \iota_N \neq 0. \tag{24}$$

Since $N < Z$, it follows that we can rewrite the first condition in Equation (25) as:

$$\mathcal{M}^T \mathcal{M}\vec{y} = \vec{0}_N. \tag{25}$$

Provided that $N \times N$ matrix $\mathcal{M}^T\mathcal{M}$ is non-singular, the only solution to (25) is the zero vector. We therefore require that the second moment matrix of state prices is full rank.

Step 2: Equivalence with the Market Equilibrium:

We next verify that the fictitious competitive economy constitutes a competitive equilibrium with no trade.

Note that, by construction, the fringe has the same consumption in both economies. The fringe's state prices are therefore the same in both economies (i.e., $q(z)$). For the fringe to correctly price all assets, we require that:

$$\sum_{z\in\mathcal{Z}} \tilde{x}_m(z)\, q(z) = \iota_M \forall m \in \{1, ..., M\}. \tag{26}$$

Since $\tilde{x}$ is a return matrix (dividend divided by price), we have the freedom to rewrite $\tilde{x}_m(z)$ as $x_m(z)/p_m$ for security price p_m. Consequently, we can use Equation (26) to find price p_m and dividend processes $x_m(z)$ that satisfy no arbitrage for the fringe. Consequently, asset prices p_m for $m \in \{1, ..., M\}$ satisfy:

$$p_m = \sum_{z\in\mathcal{Z}} x_m(z)\, q(z)\,. \tag{27}$$

Notice next that the fictitious competitive economy is constructed to satisfy the Euler Equations of strategic agents at no trade when they behave competitively. Since we endow these agents with their consumption allocations in the market equilibrium, the equilibrium consumption allocations for strategic agents are also the same in both economies.

Finally, we must check whether the consumption allocation has the same value in the fictitious competitive economy (can be financed with each agent's initial resources).

For the allocation to have the same value, the restricted asset span portfolio must have the same cost as the complete markets portfolio for all large agents:

$$\sum_{z\in\mathcal{Z}} q\left(z\right) a_i\left(z\right) = \sum_{m\in\{1,\ldots,M\}} p_m\left(c_{i2}\left(z\right) - y_i\left(z\right) k_i - s_i\right). \tag{28}$$

Substituting 27 into 29, this condition reduces to:

$$\sum_{z\in\mathcal{Z}} q\left(z\right)\left(a_i\left(z\right) - \sum_{m\in\{1,\ldots,M\}} x_m\left(z\right)\left(c_{i2}\left(z\right) - e_i\left(z\right)\right)\right) = 0, \tag{29}$$

where we can switch the order of summations because Z is finite and $M \leq Z$. Notice, however, that:

$$a_i\left(z\right) - \sum_{m\in\{1,\ldots,M\}} x_m\left(z\right)\left(c_{i2}\left(z\right) - e_i\left(z\right)\right) = 0\ \forall\, z \in \mathcal{Z}, \tag{30}$$

holds trivially by definition for the allocations to be replicated in the fictitious economy. Consequently, the equilibrium allocation has the same value, consistent with no arbitrage for strategic agents' consumption portfolios.

Step 3: Orthogonality to the Market-Implied Stochastic Discount Factor:

Consider the Hansen and Jagannathan (1991) decomposition of an admissible stochastic discount factor (SDF), $\lambda_i\left(z\right) = \Lambda_i\left(z\right)/\pi\left(z\right)$, into:

$$\lambda_i\left(z\right) = \lambda\left(z\right) + \left(\lambda_i\left(z\right) - \lambda\left(z\right)\right), \tag{31}$$

where a generic $\Lambda\left(z\right)$ is a state price deflator. This deflator is implied by market prices if:

$$\lambda\left(z\right) = \bar{\Lambda} + \left(\tilde{x}\left(z\right) - \tilde{x}\Pi\right)\beta\left(\bar{\Lambda}\right), \tag{32}$$

and $\bar{\Lambda}$ is the market-implied mean of the state price deflator, $\bar{\Lambda} = \left[\sum_{z\in\mathcal{Z}} q\left(z\right)\right]$, or the inverse of the market-implied riskfree rate, and Π is the $|\mathcal{Z}| \times 1$ vector with entries $\pi\left(z\right)$. As any admissible SDF, $\lambda\left(z\right)$, necessarily takes this functional form, so does the unique market-implied SDF, $\lambda^*\left(z\right)$, which is also the minimum variance SDF for a given SDF mean.

From Hansen and Jagannathan (1991), $\lambda^*\left(z\right)$ is defined by:

$$\beta\left(\bar{\Lambda}\right) = \Sigma^{-1}\left(\iota_M - \tilde{x}\Pi\bar{\Lambda}\right), \tag{33}$$

where Σ is the covariance matrix of returns. By construction, one has that:

$$Cov\left(\lambda^{*}\left(z\right),\lambda_{i}\left(z\right)-\lambda^{*}\left(z\right)\right)=0. \tag{34}$$

For any $\lambda_i(z)$ and $\lambda_{i'}(z)$, by the linearity of the covariance operator we have:

$$Cov\left(\lambda^{*}\left(z\right),\lambda_{i}\left(z\right)-\lambda_{i'}\left(z\right)\right)=0. \tag{35}$$

Substituting with Equation (33) into Equation (32):

$$\lambda^{*}=\bar{\Lambda}+\left(\tilde{x}-\Pi\tilde{x}\right)\Sigma^{-1}\left(\iota_{M}-\Pi\tilde{x}\bar{\Lambda}\right), \tag{36}$$

and it follows that for each security m:

$$Cov\left(\tilde{x}_{m},\lambda_{i}\left(z\right)-\lambda_{i'}\left(z\right)\right)=0\ \forall\ m,\ i,\ i', \tag{37}$$

which reveals that the asset span is orthogonal to all residual gains from trade among strategic agents.

A.3 Proof of Proposition 4

Step 1: Construction of Bespoke Economy with Restricted Asset Span:

First, consider an arbitrary payoff vector X with M assets and prices p. Suppose we can construct asset positions for each type of agent, a^I_{im}, for $i \in \{1, ..., N\}$ and $m \in \{1, ..., M\}$, that replicate the consumption allocations in complete markets without any trading restrictions:

$$X^T a^I_i = a^C_i\ \forall\, i \in \{1, ..., N\}, \tag{38}$$

where a^I_i and a^C_i are the vectors of asset positions for the restricted and complete markets, respectively.

Second, because all consumption allocations at date 2 are the same, this also applies to the fringe. State prices must therefore be the same under both market structures, $\Lambda^I_f = \Lambda^C_f$. Since the fringe's state prices are the Arrow prices with complete markets, q, and there is no arbitrage, asset prices in the restricted asset span satisfy:

$$p = Xq = X\Lambda^C_f. \tag{39}$$

Third, we establish consumption allocations are the same at date 1, i.e., that the complete markets consumption allocations are marketed with the restricted asset span. Given (38) and (39), however, this result is immediate since:

$$p^T a_i^I = q^T X^T a_i^I = q^T a_i^C \ \forall\, i \in \{1, ..., N\}.$$

Therefore, the complete markets equilibrium allocations are marketed.

Fourth, we establish all agents' decision problems are satisfied. The fringe's Euler Equations are trivially satisfied because asset prices are derived from its state prices. For strategic agent i, we can rewrite its first-order necessary condition for its optimal asset demand as:

$$p + \mu_i \Gamma^I a_i^I = X \Lambda_i^I, \tag{40}$$

where Γ^I is the price impact matrix under the restricted asset span. By no arbitrage from (39), we recognize:

$$\Gamma^I = X \Gamma^C X^T, \tag{41}$$

where Γ^C is the diagonal complete markets price impact matrix with diagonal entries $\frac{\partial q(z)}{\partial A(z)}$. It is straightforward to see this matrix is symmetric and full rank.

Substituting these results into (40), and recognizing $\Lambda_i^I = \Lambda_i^C$ because equilibrium allocations are the same, the first-order necessary conditions then reduce to:

$$a_i^I = \mu_i^{-1} \left(X \Gamma^C X^T \right)^{-1} X \left(\Lambda_i^C - \Lambda_f^C \right). \tag{42}$$

Notice when $X = Id_Z$, so that markets are complete, then:

$$a_i^C = \mu_i^{-1} \left(\Gamma^C \right)^{-1} \left(\Lambda_i^C - \Lambda_f^C \right). \tag{43}$$

Substituting (43) into (42), we arrive at:

$$a_i^I = \left(X \Gamma^C X^T \right)^{-1} X \Gamma^C a_i^C. \tag{44}$$

Substituting (38) into (44), we arrive at the identity $a_i^I = a_i^I$.[10] Consequently, the first-order necessary conditions of strategic agents are also satisfied.

[10]Notice we can derive (44) from (38) by using the price impact matrix, $\left(\Gamma^C\right)^{-1}$, as a weighting matrix to project X onto a_i^C.

Step 2: Welfare Comparison to Complete Markets:

Finally, we establish that we can find a restricted asset span (rank deficient X) that can replicate the complete markets allocation and consequently deliver the same welfare. Let $\tilde{\Lambda}_i = \left(\frac{\mu_i}{m_f}\right)^{-1}(\Lambda_i - \Lambda_f)$, and $\tilde{\Lambda}$ the $N \times Z$ matrix that stacks these state prices. Now conjecture $X = \tilde{\Lambda}\left(\Gamma^C\right)^{-1}$, and let a^I bet the $M \times N$ matrix of strategic agent asset positions, and a^C its complete markets counterpart. Stacking (43) across agents into a matrix equation and substituting this conjecture, we arrive at:

$$X^T a^I = X^T\left(X\Gamma^C X^T\right)^{-1} X\tilde{\Lambda}^T = \left(\Gamma^C\right)^{-1}\tilde{\Lambda}^T = a^C.$$

Further, if $N < Z$, then X, which is a $N \times Z$ matrix, is rank deficient. Consequently X is the payoff matrix for a restricted asset span.

We next establish the restricted asset span delivers higher welfare than complete markets.[11] The Planners' objective is

$$max_X V = \sum_i u_{i2}(a_i' X(z) + y_i(z)) + u_f(a_f' X(z) + y_f(z)) + \sum_i u_{i1}(w_i - a_i' q) - a_f' q$$

subject to the first-order necessary conditions of all agents. The first-order necessary conditions are $q = X\Lambda_f$ for the competitive fringe and $a_i = \mu^{-1}(X\Gamma X')^{-1}X\tilde{\Lambda}_i$ for strategic agents of type i, where $\Gamma = \text{diag}(\pi(z)|u_f''(c_f(z))|)$ and $\tilde{\Lambda}_i = \Lambda_i - \Lambda_f$ can be written as $X\Gamma X' a_i = \mu^{-1}X\tilde{\Lambda}_i$. The j^{th} row of this equation is $X_j\Gamma X' a_i = \mu^{-1}X_j\tilde{\Lambda}_i$, where X_j is j^{th} asset. Suppose λ_{ij} is the Lagrange multiplier associated with this constraint. As such, $\frac{\partial X_j \Gamma X' a_i}{\partial X_s(z)} = \frac{\partial \mu^{-1} X_j \tilde{\Lambda}_i}{\partial X_s(z)}$.

Consider now the economy with the restricted asset span that replicates the complete markets allocation. In this economy, $a_i = e_i$, where e_i is the vector with the i^{th} entry equal to one and the rest are zeros. As a result,

$$\frac{\partial X_j \Gamma X_i'}{\partial X_s(z)} = \mu^{-1} 1_{\{j=s\}}\tilde{\Lambda}_i(z) + \mu^{-1}X_j(z)\frac{\partial \Lambda_i(z)}{\partial X_s(z)} - \mu^{-1}X_j(z)\frac{\partial \Lambda_f(z)}{\partial X_s(z)},$$

which can be written as $F(j,i,s) = \mu^{-1}X_j(z)\Theta_i(z)1_{\{i=s\}}$, where

$$F(j,i,s) = \frac{\partial X_j \Gamma X_i'}{\partial X_s(z)} - \mu^{-1}1_{\{j=s\}}\tilde{\Lambda}_i(z) + \mu^{-1}X_j(z)\frac{\partial \Lambda_f(z)}{\partial X_s(z)}.$$

[11]Formally if model parameters have a multivariate continuous valued distribution then the probability complete market is optimal is zero

The system of equations can be rewritten in matrix form as

$$\left[\frac{\partial V}{\partial X_s(z)}\right]_{sz,1} = [\mu^{-1}X_j(z)\Theta_i(z)1_{\{i=s\}} - F(j,i,s)]_{sz,ji}[\lambda_{ji}]_{ji,1},$$

with $\Theta_i(z) = u_i''(c_i(z))$. This can be written more concisely as

$$v_{NZ\times 1} = M_{NZ\times N^2}\lambda_{N^2\times 1}.$$

Suppose the first rows are split into N^2 and $N(Z-N)$ row blocks. This results in two systems of equations $v_1 = M_1\lambda$ and $v_2 = M_2\lambda$. As such, $\lambda = (M_1)^{-1}v_1$. As a result $v_2 = M_2(M_1)^{-1}v_1$, or $v_2 = M_2v_3$ where $v_3 = (M_1)^{-1}v_1$. Note that v_2, M_1, and v_1 do not contain $\Theta_i(z)$, which is the second derivative of strategic agents' utility. Therefore, each row of the equation is equivalent to a linear relation between $\Theta_i(z)$, with coefficients including $\Lambda_i(z), \Gamma(z)$, and $u'''(c_f(z)), u_{2i}'(c_i(z))$, and $u_{1i}'(c_i)$.

The extra equations guarantee that the probability that complete markets is optimal is zero. The only other case for this equation to hold generically is the coefficient for $\Theta_i(z)$ is zero. This means that $X(z)'\lambda_i = 0$ for all z, i. Both vectors are of size N . Fix i. Notice that $X_1, X_2, ..., X_N$ are perpendicular to λ_i; as a result, $X = [X_1X_2...X_N]'$ must be rank deficient, which is not true. To see this, we recognize that for complete markets, $a_i = \mu^{-1}\Gamma^{-1}\tilde{\Lambda}_i$, and therefore $X = \Gamma^{-1}\tilde{\Lambda}$. As such, for X to be rank deficient, $\tilde{\Lambda}$ must also be rank deficient, which is not true generically. The only remaining case is when $\lambda_i = 0\ \forall i$. In this case, $\frac{\partial V}{\partial X_s(z)} = 0\ \forall\ s, z$. Since $q = X\Lambda_f$, $\frac{\partial q_j}{\partial X_i(z)} = \Lambda_f(z)\delta\{i=j\}$. Plugging this into the welfare objective, V, leads to

$$u_{2i}'(c_i(z))\pi(z) - u_{2f}'(c_f(z))\pi(z) - u_{i1}'(w_i - a_i'q)\Lambda_f(z) + \Lambda_f(z) = 0, \forall i, z.$$

This implies $\forall i, z : \tilde{\Lambda}_i(z) = 0$. If this is true, then $\forall i : a_i = 0$ and autarky would be optimal, which is not the case.

A.4 Proof of Proposition 5

Step 1: Pure Idiosyncratic Risk Case:

First, consider the case of complete markets without any restrictions on what agents can trade. In this case, the asset price is the same for both the high and low states by symmetry. Let it be q_{cmp}. Let $a_B > 0$ be the position of an agent when it buys the asset and $a_S < 0$

be the position when it sells. The utilitarian welfare is:

$$\begin{aligned}\mathcal{W}(I_2,\mu) &= 2u\left(\bar{y}-q_{cmp}\left(a_B+a_S\right)\right)+u\left(\bar{y}-\Delta+a_B\right)+u\left(\bar{y}+\Delta+a_S\right)\\ &\quad +2q_{cmp}\left(a_B+a_S\right)+m_f u\left(\bar{y}-\frac{a_B+a_S}{m_f}\right).\end{aligned}$$

Recall the fringe has the same time 2 preferences as the strategic agents. As a result of the convexity of marginal utility with symmetric preferences, sellers restrict their asset positions more than buyers. As such, $a_B+a_S>0$.

In contrast, with the aggregate swap, both agents take a symmetric position of a when buying and $-a$ when selling, and internally clear the swap market without the fringe. Utilitarian welfare is then:

$$\mathcal{W}([1-1],\mu)=2u\left(\bar{y}\right)+u\left(\bar{y}-\Delta+a\right)+u\left(\bar{y}+\Delta-a\right)+m_f u\left(\bar{y}\right).$$

Since price impact is convex because it is the fringe's marginal utility, it follows that:

$$q_{cmp}=u'\left(\bar{y}-\left(a_B+a_S\right)\right)>u'\left(\bar{y}\right)=q_{inc}.$$

As such, $|a_S|<a_B<a$ because insurance is more expensive in complete markets.

Let $\Delta_a=\Delta-a\geq 0$, $\delta_B=a_B-a\leq 0$, and $\delta_S=a_S+a\geq 0$. Because $|a_S|<a_B$, it follows that $\delta_B+\delta_S=a_B+a_S>0$. Since $u\left(\cdot\right)$ is strictly concave, we have that:

$$\begin{aligned}u\left(\bar{y}-\Delta_a\right)+u\left(\bar{y}+\Delta_a\right)+m_f u\left(\bar{y}\right) &\geq u\left(\bar{y}-\Delta_a+\delta_B\right)+u\left(\bar{y}+\Delta_a+\delta_S\right)\\ &\quad +m_f u\left(\bar{y}-\frac{\delta_B+\delta_S}{m_f}\right),\end{aligned} \tag{45}$$

with equality when $\mu=0$. This is because the arguments in the left-hand and right-hand sides of (45) are just a reshuffling of the allocations:

$$\bar{y}-\Delta_a+\bar{y}+\Delta_a+m_f\bar{y}=\bar{y}-\Delta_a+\delta_B+\bar{y}+\Delta_a+\delta_S+m_f\bar{y}-\delta_B-\delta_S,$$

in which the fringe consumes less and each strategic agent has higher volatility in their consumption.

Consequently, it follows:

$$\mathcal{W}(I_2,\mu) \leq 2\left(u\left(\bar{y}-q_{cmp}\left(a_B+a_S\right)\right)-u\left(\bar{y}\right)+q_{cmp}\left(a_B+a_S\right)\right)+\mathcal{W}([1-1],\mu).$$

Suppose that $u'(\bar{y}) \geq 1$. Then it is immediate from a 1st-order Taylor expansion of $u\left(\bar{y} - q_{cmp}\left(a_B + a_S\right)\right)$ around $\bar{y}$ that:

$$\mathcal{W}(I_2, \mu) \quad < \quad 2\left(1 - u'\left(\bar{y}\right)\right) q_{cmp}\left(a_B + a_S\right) + \mathcal{W}([1-1], \mu) < \mathcal{W}([1-1], \mu),$$

because the 2nd-order term is negative.

Step 2: Pure Aggregate Risk Case:

Let $x_1 = [a_1^{cmp}(L), a_1^{cmp}(H)]$ be a payoff vector based on the asset demands of the less risk averse agent (agent 1) in the complete markets competitive equilibrium with $a_1^{cmp}(L)a_1^{cmp}(H) < 0$ and no restrictions on what agents can trade. In addition, let $x_f = [a_f^{cmp}(L), a_f^{cmp}(H)]$ be a payoff vector based on the asset demands of the competitive fringe in the competitive equilibrium.

We now consider the following thought experiment for the competitive fringe. Suppose the fringe trades an asset x that can be different from what strategic agents trade. By varying x between x_1 and x_f, we can continuously vary the trading positions of the fringe between the restricted asset span and the complete markets economies.

The first-order necessary condition of strategic agent i in an economy with a restricted asset span with an asset whose payoff is x is:

$$\sum_z x(z)(\Lambda_i(z) - \Lambda_f(z)) = (3-2i)a_i \frac{\mu}{m_f} \sum_z X^2(z)|u_f''(e(z) + a_f X(z))|,$$

where we recall that the asset price is $q_1 = \sum_z x(z)\Lambda_f(z)$ and $i \in \{1,2\}$.

We can then sum the first-order necessary conditions over the two strategic agents to arrive at:

$$\sum_{z,i} x(z)\Lambda_i(z) = 2\sum_z x(z)u_f'(e(z) + a_f x(z)), \tag{46}$$

where we assume m_f is small so that $a_1 + a_2 \approx 0$. Suppose (with abuse of notation) $a_f = G_f(a_1; X)$ is the solution to the following equation to (46). If $X = X_1$, then the allocations will be the same as without any asset-trading restrictions.

Also suppose that $a_1 = G_1(a_f; X)$ is the solution to:

$$\sum_{z,i}(3-2i)X_1(z)\Lambda_i(z) = 2a\frac{\mu}{m_f}\sum_z {X_1}^2(z)|u_f''(e(z) + a_f X(z))| \tag{47}$$

where a is asset supply by agent 1, and we derived (47) by subtracting the first-order necessary conditions of the two types of strategic agents.

Finally, we define $f(a;X) = G_1(G_f(a;X);X)$. Note that the solution to the restricted asset span problem is equivalent to finding a fixed point of the function $f(a;X_1)$ and the solution to the complete markets problem is a fixed point of $f(a,X_f)$.

Suppose $\mu > 0$. Then, with complete markets agent 1 under-supplies the Arrow-Debreu asset referencing state L state. To see this, from the first-order necessary conditions of agent 1, $a_1(L) < 0 < a_1(H)$ and $u_1'(e(L) - |a_1(L)|) < u_f'(e(L) - |a_f(L)|)$. As a result, $e(L) - |a_1(L)| > e(L) - |a_f(L)|$, and consequently $|a_f(L)| > |a_1(L)|$. Similarly, one has that $u_1'(e(H) + |a_1(L)|) > u_f'(e(H) + |a_f(H)|)$. As a result, $e(H) - |a_1(H)| < e(H) - |a_f(H)|$, and therefore $|a_1(H)| > |a_f(H)|$ It then follows from these observations that $|\frac{a_f^{cmp}(L)}{a_f^{cmp}(H)}| > |\frac{a_1^{cmp}(L)}{a_1^{cmp}(H)}|$.

This reveals the fringe supplies too much in state L when markets are complete. We next show that the restricted asset span leads to lower market power for agent 1.

To show this, consider an arbitrary $\theta \leq 1$. For a given a, we can take the derivative of equation 46 with respect to θ for $a_f(\theta)$, one has:

$$0 = X_1(L)X(L)u_f''(e(L) + a_f X(L))\frac{\partial a_f}{\partial \theta} + X(H)X_1(H)u_f''(e(H) + \theta a_f X_1(H))\frac{\partial \theta a_f}{\partial \theta}.$$

By above equation, one can conclude that $\frac{\partial a_f}{\partial \theta} < 0 < \frac{\partial \theta a_f}{\partial \theta}$. It then follows that:

$$\frac{\partial \sum_z X_1(z)|u_f''(e(z) + a_f X(z))|}{\partial a_f} = -\sum_z X^2(z)X_1(z)u^{(3)}(c_f(z))\frac{\partial \Theta(z)a_f}{\partial \theta} < 0,$$

where $\Theta(L) = 1, \Theta(H) = \theta$. Therefore, as θ increases, price impact decreases, and implied asset supply by the less risk averse agent from equation (47) consequently increases.

As a result, $f(a;[X_1(L), X_1(H)]) > f(a;[X_1(L), X_1(H)\theta_0])$ for $\theta_0 < 1$. By setting $\theta_0 = \frac{X_f(H)X_1(L)}{X_f(L)X_1(H)}$, we conclude that $f(a;X_1) > f(a;X_f)$. This establishes that total price impact will be lower with the restricted asset span, and the implied a of agent 1 will be higher.

Given that $a^{cmp} = f(a^{cmp};X_f)$, for $0 \leq a \leq a^{cmp}$ one has $f(a;X_f) > a$ and hence the fixed point (which is a^{incmp}) of the function $f(a,X_1)$ cannot be in $[0,a^{cmp}]$. Therefore, it must be the case that $a^{incmp} > a^{cmp}$.

What remains to be shown is that welfare is increasing in the asset supply of the less risk averse agent. It is immediate that:

$$\frac{\partial W}{\partial a} = \overbrace{X_L(\Lambda_1(L) - \Lambda_{-1}(L))}^{\geq 0} + \overbrace{X_H(\Lambda_1(H) - \Lambda_{-1}(H))}^{\geq 0} \geq 0,$$

which completes the proof.