A Trilemma for Asset Demand Estimation*

Most recent version

William Fuchs† Satoshi Fukuda‡ Daniel Neuhann§

January 15, 2026

Abstract

This paper derives fundamental limits to identifying asset demand from observational data. We establish a trilemma: it is impossible to maintain that (i) prices satisfy no arbitrage, (ii) investors value assets for their payoffs, and (iii) asset demand curves are invariant to exogenous asset supply shocks. That is, one cannot use supply shocks to move along an asset demand curve without shifting it. The only exception is the knife-edge case in which the asset menu consists of Arrow securities. In realistic settings, demand elasticities thus necessarily reflect theoretical assumptions rather than the data alone.

*We thank Dan Ackerberg, Yu An, Olivier Darmouni, Aureo de Paula, Julie Fu, Andrey Ordin, and seminar participants at UCL, UT Austin, University of Zurich, 7th Columbia Workshop in New Empirical Finance, Tepper-LAEF Advances in Macrofinance Conference, and WashU Finance Conference for helpful comments. Xinran Zhang provided outstanding research assistance.

†University of Texas at Austin, FTG, and CEPR. william.fuchs@mccombs.utexas.edu, https://orcid.org/0000-0001-9936-6712

‡Santa Clara University. sfukuda@scu.edu.

§University of Texas at Austin and FTG. daniel.neuhann@mccombs.utexas.edu.

1 Introduction

Estimating demand functions using supply shocks is a cornerstone of empirical economics. Can this approach be fruitfully applied to financial assets? We show that demand estimation in asset markets is highly constrained by two fundamental principles of asset pricing: (i) prices admit no arbitrage, and (ii) asset demand is, at least in part, *instrumental*: preferences are defined over assets' payoffs, not asset holdings directly. Under these principles, even perfectly exogenous asset supply shocks generically induce *shifts* in the very demand curve that one wants to analyze. This prevents the non-parametric estimation of asset demand elasticities.

The basic principle of demand estimation with supply shocks is to generate as-if-random variation in a given price, holding all other relevant variables fixed. The central problem in financial markets is that assets are bundles of state-contingent payoffs, and that these can be flexibly *recombined and unbundled* through portfolio formation. This forces the price system to satisfy the consistency requirement of no arbitrage: to ensure well-defined asset demand functions, each asset's price must correspond to the value of its underlying payoffs. Any supply-induced change in an asset price must therefore produce consistent changes in the prices of the underlying payoffs as well. But if no arbitrage is to hold, the prices of all assets which pay off in at least some overlapping states of the world must *also* change. This is a failure of the *ceteris paribus* condition.

This failure alone need not threaten identification: if the demand system is separable across assets, correlated price changes do not contaminate the demand response. However, portfolio choice is known to feature *endogenous demand complementarities*: the marginal value of any asset depends on its covariance with the rest of the investors' portfolio. Asset demand functions thus generically depend on the prices of all assetse in the choice set, and correlated price changes *shift* the very demand curve one wanted to measure.

We summarize these concerns as a trilemma: one cannot maintain that (i) prices satisfy no arbitrage, (ii) investors have preferences over payoffs, and (iii) supply shocks can be used to move along a demand curve without shifting it.

These challenges are a distinct feature of financial markets. In conventional goods markets, product characteristics are determined by suppliers and cannot be autonomously reconfigured by buyers. Hence these markets thus feature no cross-price restrictions akin to no arbitrage – the price of a car can deviate from the cost of its components without inducing sudden trade in car parts. This permits a stable dichotomy between supply and demand that is untenable in financial markets.

To see our result more formally, write asset prices in terms of the *state prices*. Under preferences over payoffs, they measure what investors care about: the cost of a unit state-contingent cash flow. Let p be the $J \times 1$ vector of asset prices, Y the $J \times Z$ payoff matrix summarizing the state-specific payoffs of J assets, and q the $Z \times 1$ vector of state prices. No arbitrage implies that asset prices satisfy

$$p = Yq.$$

The *ideal experiment* underlying demand estimation considers an exogenous shock to a single asset price. This experiment can be stated in terms of state prices. To allow incomplete markets, let Y^+ denote the Moore-Penrose pseudo-inverse. Then $q = Y^+ p$ and an asset price shock affects state prices according to

$$\frac{\partial q}{\partial p^{\mathrm{T}}} = Y^+.$$

That is, the ideal experiment requires a specific change to the state price vector which is fully determined by the *inverse* payoff matrix.

In contrast, the state price response to a supply shock reflects the change in marginal value of state-contingent payoffs. Since this reflects the total supply of state-contingent payoffs, the effects of the supply shock are thus proportional to the payoff matrix Y itself, *not its inverse*. Since a matrix and its inverse generically differ, a supply shock generically fails to reproduce the ideal experiment.

What is worse, the payoff matrix and its inverse generically have elements of the *opposite sign*. This means that a supply shock makes certain payoffs cheaper when the ideal experiment requires them to become more expensive. Such directional errors naturally create large biases when estimating substitution patterns.

The only exceptions to this result are knife edge. In complete markets, the menu would need to consist of Arrow securities, so that the payoff matrix is diagonal. In incomplete markets, the analogous condition is that there are no assets with overlapping payoffs: *for each state, there is at most one asset with a positive payoff*.

Real-world assets are far from satisfying these restrictions. While this is intuitively clear, we further substantiate this fact in two ways. First, we study random payoff matrices drawn from distributions with factor structures. We find that approximately *half* of the elements of the inverse payoff matrix have the wrong sign. More disconcertingly, the chance that any given element has the wrong sign is a coin flip. This instability makes it difficult to systematically control for potential errors. Second, we randomly draw (subsets) of payoff matrices from stocks in the S&P 500 and find the same broad patterns. Taken together, there is little reason to suspect that the barriers to identification we derive are immaterial in practice.

Neither no arbitrage nor payoff-based preferences are easily discarded. Preferences over payoffs form the basis of portfolio choice theory. Given such preferences, no arbitrage ensures the existence of smooth demand functions that do not jump discontinuously in response to small price changes—as is required for consistent demand estimation and counterfactuals. No arbitrage is also a weak equilibrium requirement which is likely to hold at least approximately in observational data: even if trading is frictional and prices temporarily admit (near) arbitrage, such frictions generally do not eliminate cross-asset price linkages and longer run portfolio holdings naturally depend on some common pricing kernel. Thus, the tendency for asset prices to reflect common state prices remains even in these settings. Precisely for this reason, existing asset demand estimation approaches rely on no arbitrage to derive tractable demand systems based on a small number of characteristics and risk factors (Koijen and Yogo, 2019).

In principle, it is possible to overcome the identification issues we document using sufficiently many independent shocks. This echoes classic work in demand identification in Mas-Colell (1977). To investigate the necessary data requirements, we consider an idealized setting where the econometrician observes N

independent "experiments," each of which generates price and quantity variation. Can we reliably identify asset-level demand elasticities from this data? While asset demand curves are naturally non-linear, we consider the favorable benchmark with linear demand curves. With J assets, each investor's portfolio choice problem is then characterized by a $J \times J$ cross-substitution matrix. Hence one can point-identify the entire matrix of demand parameters *only if* N is no smaller than J, the dimensionality of the asset span. With fewer than J experiments, identification collapses to projections onto the subspace of observed price changes, and demand parameters are arbitrarily unconstrained outside the span of observed shocks.

These data requirements are stringent. In many financial markets, the number of assets available to trade is very large—potentially in the hundreds or thousands. However, even if one considers only a handful of aggregate portfolios such as bond and stock portfolios, one needs a setting with multiple shocks that produce independent variation. Since many financial markets are connected (and thus do not provide independent variation), a natural candidate is variation over time. Here one faces the problem that the pricing kernel (i.e., the stochastic discount factor) has a *permanent component* (Alvarez and Jermann, 2005; Borovička, Hansen, and Scheinkman, 2016). Such lack of stationarity implies that observing shocks over time is unlikely to provide the required independent variation.

Taken together, our results show that structural models or other theoretical restrictions are necessary to recover demand elasticities from observational data on portfolio holdings and prices. As such, estimated demand elasticities must be understood, at least in part, as theoretical objects determined by a priori assumptions, not by empirical evidence. More constructively, we show that characterizing the divergence between ideal experiment and supply shocks in terms of state prices provides insights into sources of misspecification, and how assumptions on preferences and portfolio construction can be used to mitigate bias.

Related literature. Our paper relates to an important literature in finance and economics studying demand effects in financial markets. Early work in this area includes portfolio balance models (Tobin, 1969), and the price effects of index inclusions in equity markets (Shleifer, 1986; Harris and Gurel, 1986). More recently, this broad mechanism has found applications in unconventional monetary policy, foreign exchange markets, and fund flows in bond and equity markets.

This rightly influential literature shows that constraints on capital flows can have important effects on asset prices. However, it stops short of systematically establishing whether and when these price effects reveal structural aspects of investor and market behavior. This is important because critical aspects of asset price determination and policy transmission tightly depend on the price responsiveness of financial markets. We find that non-parametric approaches generically fail to identify asset demand elasticities because they are contaminated by cross-price effects. This means that implicit or explicit theoretical restrictions play a central role in determining the interpretation and policy relevance of the documented effects.

One consequence of our findings is structural methods are important tools for understanding demand effects in asset markets, much like in many other settings (Berry and Haile, 2021). However, asset markets present particular challenges: preferences are instrumental, investors can autonomously reconfigure "products," and choice is continuous. This means that one cannot easily turn a decision problem with complementarities into, e.g., a discrete-choice problem over bundles. These differences clarify our relationship to recent work in industrial organization which estimates demand systems with complementarities (e.g., Iaria and Wang, 2020; Wang, 2024; Fosgerau, Monardo, and de Palma, 2024; Ershov, Laliberté, Marcoux, and Orr, 2024). These approaches typically study settings in which consumers make discrete choices over a limited number of bundles, or where substitution patterns are governed by exogenous functional-form parameters. As such, the source of complementarities and the methods to deal with them are distinct from no arbitrage and portfolio spillovers. The spirit of the exercise is also different. While these papers develop tools for estimating demand with complementarities under exogenous price variation, we asks whether such demand can be identified from supply shocks under stringent cross-price restrictions.

To overcome these challenges, structural models of asset demand must account for the cross-asset linkages and price spillovers inherent to portfolio choice.

Fuchs, Fukuda, and Neuhann (2025a) show that the prominent logit approach in Koijen and Yogo (2019) generally fails to do so and that this can lead to large biases in estimated demand elasticities. While we focus on a static setting, the same issues would also arise in a dynamic setting where investors can trade securities referencing different states and dates, as these would also have to be priced by a common pricing kernel and governed by no arbitrage. This broader view helps connect our findings to those in Binsbergen, David, and Opp (2025) and He, Kondor, and Li (2025). Allen, Kastl, and Wittwer (2025) propose a structural model to estimate asset demand without reliance on price instruments. Consistent with our results, this approach requires a-priori restrictions and uses data on bid schedules.

Another approach is to circumvent the issue of cross-asset spillovers by estimating *differences* in demand curves across similar assets subject to symmetric spillovers (Haddad, He, Huebner, Kondor, and Loualiche, 2025). While this restores identification under specific assumptions on the endogenous substitution matrix, it comes at the cost of identifying individual demand curves. Haddad, He, Huebner, Kondor, and Loualiche (2025) further argue that, under their strict assumptions, one can identify the entire substitution matrix using time-series variation and shocks to factor portfolios. As Section 7 discusses, this requires as many independent experiments as the dimensionality of the asset span. Such rich data is difficult to obtain when the pricing kernel has a permanent component. Hence the limits to non-parametric identification we derive also apply to this methodology.

An important special case is an asset menu with Arrow securities. While typical assets are not Arrow securities, one can attempt to construct them from other assets. Unfortunately, this requires knowing the unobserved payoff matrix. An (2025) and An and Huber (2024) pursue a related approach by constructing portfolios orthogonalized with respect to returns and flows to specific investors. However, orthogonal payoffs are not sufficient to ensure no overlap in the payoff distribution. Their approach thus requires additional assumptions, including that demand for uncorrelated portfolios is independent of each other. Consistent with our results, this is an a-priori restriction on substitution patterns.

2 Setup

We consider a canonical model with a set I of potentially heterogeneous investors. Each investor $i \in I$ must choose how much to consume at date 0 and across Z states of the world at date 1. To acquire a desired state-contingent consumption profile, the investor can invest in J assets. Our definition of assets is entirely generic and covers "primitive" assets, such as stocks, and portfolios composed of other assets.

Investor i's *portfolio* is a vector $a^i \equiv (a^i_j)^J_{j=1} \in \mathbb{R}^J$ of asset positions, where each element a^i_j denotes the investor's holdings of asset j. Asset j has payoff $y_j(z)$ in state z. The probability of state z is $\pi_z \in (0,1)$. We denote by $Y \equiv (y_j(z))_{j,z}$ the $J \times Z$ matrix of cash flows. The payoff matrix is known to the investor but unobserved by the econometrician. This is because payoffs reflect expected returns, which are latent. Prices are observed by both the investor and the econometrician.

We treat time-zero consumption as the numeraire good (or, equivalently, as the *outside asset*) whose price is normalized to 1. Investor i is endowed with $e^i_j \geq 0$ units of asset j and $e^i_0 \geq 0$ units of the numeraire. Each investor also receives unobserved non-traded endowments w^i_0 and $w^i(z)$ at date 0 and in state z, respectively.

Decision problem. Investor i has strictly increasing and strictly concave von-Neumann Morgenstern utility function u^i defined over state-contingent consumption and discount factor $\delta^i \in (0,1)$. Beyond the standard budget constraints, each investor may face other constraints on portfolio formation. Let $\mathcal{A}^i$ denote the set of feasible portfolios of investor i, and assume that $\mathcal{A}^i$ is a closed convex subset of $\mathbb{R}^J$. The decision problem is

$$\sup_{a^i \in \mathcal{A}^i} \quad (1-\delta^i)u^i(c^i_0) + \delta^i \sum_{z=1}^{Z} \pi_z u^i(c^i_z) \tag{1}$$

$$\text{s.t} \quad c^i_0 = e^i_0 - \sum_{j=1}^{J} p_j(a^i_j - e^i_j) + w^i_0 \quad \text{and}$$

$$c^i_z = \sum_{j=1}^{J} y_j(z) a^i_j + w^i(z) \quad \text{for all } z.$$

Solution and Demand System. We stress two properties of this decision problem. First, as the next section discusses in detail, the existence of a smooth, well-defined solution to the portfolio problem generally requires the absence of arbitrage. In this sense, no arbitrage is a *precondition* for well-conditioned demand analysis in asset markets. This is the first leg of the trilemma.

Second, when a solution exists, it generically consists of J asset-level demand functions $a_j^i(p)$, each of which depends on the *entire vector* of asset prices p. To see this explicitly, assume for now that an interior solution exists. Then the asset-level demand functions are implicitly defined by the J first-order conditions

$$\frac{\delta^i}{1-\delta^i}\sum_z \frac{\pi_z u'^{,i}(c_z^i)y_j(z)}{u'^{,i}(c_0^i)} = p_j \qquad \text{for all } j,$$

where $u'^{,i}(\cdot)$ indicates i's marginal utility. The marginal value of each asset thus depends on its contribution to total state-contingent consumption, which in turn is jointly determined by the holdings of all assets. Altering either the price or quantity of any asset thus typically changes the demand for other assets as well. At its root, this structure obtains because investors ultimately care about state-contingent payoffs, not asset holdings directly. This is the second leg of the trilemma.

Identification Problem and Demand Elasticities. The properties of the demand system are jointly influenced by a number of latent variables: the payoff matrix Y, preference parameters u^i and δ^i and latent non-traded endowments w_0^i and $w^i(z)$. One goal of identification might therefore be to estimate each of these latent parameters. The literature thus far is instead mainly concerned with the identification of demand functions themselves. In particular, asset-level demand functions can be characterized by the price elasticity of demand, which is the *partial derivative* of demand for asset j with respect to an asset price p_k, holding all other prices fixed,

$$\mathcal{E}_{j,k}^i \equiv -\frac{\partial a_j^i(p)}{\partial p_k}\frac{p_k}{a_j^i(p)}.$$

We will be concerned with identifying this object. As indicated, the main difficulty is that each demand function depends on the entire vector of asset prices. Hence identifying the demand elasticity requires exogenous variation in one asset price while other asset prices remain fixed.

Remark 1 (Preferences over non-pecuniary characteristics) *Some recent work in asset pricing emphasizes certain non-pecuniary motives for investing in specific assets (Starks, 2023). For example, socially responsible investors may hold a stock in part because they believe that the company is a good steward of the environment. While such motivations partially decouple asset valuations from cash flows, they generally do not do so entirely: even socially responsible investors may care at least in part about financial returns.*

Remark 2 (Dependence on multiple prices in special settings) *The dependence of demand functions on multiple asset prices exists even in settings that purportedly induce asset demand functions that depend only on the asset's own price. For example, Koijen and Yogo (2025, Appendix A) study a model with CARA preferences, normally distributed payoffs, a diagonal covariance matrix conditional on factors, and a risk-free asset with a fixed interest rate normalized to zero. The first two features generate linear marginal utility, and the combination with the third and fourth features yields separable asset demand functions that depend only on the excess expected return and volatility of a given asset. However, the independence of other prices is an illusion achieved by "normalizing" the risk free rate to a fixed number. Yet in equilibrium, the risk-free rate is not a parameter, it is the inverse sum of state prices and thus reflects all other asset prices. Separability of demand functions is therefore achieved by a-priori restrictions on asset prices.*

3 The Role of No Arbitrage for Demand Analysis

Demand analysis in financial markets faces two basic challenges. The first is that demand functions must be sufficiently well-behaved. For example, demand elasticities are partial derivatives of demand with respect to an asset price. Hence an elasticity can be used to describe demand only if the underlying demand functions are smooth functions of asset prices. The second is the large number of assets under consideration. In US equities markets alone, investors can choose among many *thousands* of assets, which creates a curse of dimensionality in demand estimation.

Both challenges can be addressed using the principle of no arbitrage. For example, the influential approach in Koijen and Yogo (2019) implicitly relies on the Ross (2004) arbitrage pricing theory to argue that asset demand can be summarized by a small number of asset characteristics and risk factors, leading to a low-dimensional representation. (We leave aside here the concern that several common characteristics, such as book-to-market ratios, are themselves endogenous to demand.) Conversely, arbitrage opportunities can lead to discontinuous changes in demand functions in response to arbitrarily small price changes. No arbitrage rules out such discontinuities, thereby ensuring a well-behaved demand system.

To lend credence to these statements, we briefly recapitulate the link between demand functions and no arbitrage. Since the empirical literature often emphasizes constraints on portfolio formation, we account for such frictions as well. We then establish the standard result that, under weak conditions, no arbitrage allows for an analysis of asset prices (and thus demand) using *state prices*.

We begin by defining *unbounded* arbitrage opportunities as those that can be exploited using arbitrarily large asset positions. Standard definitions of arbitrage always consider unbounded arbitrage opportunities (Duffie, 2001). Our analysis below differs only in that we also permit *bounded* arbitrages.

Definition 1 (No unbounded arbitrage for investor i) *Investor i has an unbounded arbitrage opportunity if, for any $m > 0$, there exists a portfolio $a^i \in \mathcal{A}^i$ such that either (i) $p \cdot a^i \leq 0$, $Y^{\mathrm{T}} a^i \geq 0$, and $(Y^{\mathrm{T}} a^i)_z \geq m$ for some z or (ii) $p \cdot a^i \leq -m$ and $Y^{\mathrm{T}} a^i \geq 0$. Otherwise, investor i has no unbounded arbitrage opportunity.*

Proposition 1 shows that a well-defined decision problem requires the absence of unbounded arbitrage opportunities. The simple reason is that unbounded arbitrage precludes the existence of a solution to the investor's problem. This is a well-known result based on textbook treatments (e.g, Duffie, 2001).[1]

[1]Since the material in this section is standard we relegate the proofs to the Online Appendix.

Proposition 1 (Duffie (2001): No arbitrage and the investor's problem) *If there is a solution to (1), investor i has no unbounded arbitrage opportunity. If U^i is continuous and investor i has no unbounded arbitrage opportunity, then there is a solution to (1).*

Asset prices and demand can then be analyzed using *state prices*, which measure the marginal cost of a unit of state-contingent consumption. In particular, if the union of investors' feasible sets covers the space of feasible portfolios $\mathbb{R}^J$, the absence of unbounded arbitrage implies the existence of state prices such that asset prices are payoff-weighted sums of state prices.

Lemma 1 (Existence of state prices) *If there exists a subset of investors I_0 such that every $i \in I_0$ does not have an unbounded arbitrage opportunity and $\mathbb{R}^J = \bigcup_{i \in I_0} \mathcal{A}^i$, there exist state prices $q \in \mathbb{R}^Z_{++}$ such that asset prices are payoff-weighted sums of state prices:*

$$p = Yq. \tag{2}$$

The same basic mechanism applies to *bounded arbitrage* as well, whereby investors can only exploit mispricing up to an exogenous constraint on asset positions. In particular, it remains optimal to exploit the arbitrage to the extent possible, and this can lead to discontinuous changes in demand functions in response to arbitrarily small price changes. As Example 2 in Appendix B.1.1 illustrates, this remains the case even though unbounded arbitrages are ruled out by portfolio constraints. Merely asserting the presence of portfolio constraints is thus not sufficient to have a well-posed estimation problem. Since any infinitesimal price change triggers an arbitrage for redundant assets, for the remainder we focus on the more interesting case without redundant assets.

Assumption 1 (No redundant assets) $Z \geq J$ *and* $\text{rank}(Y) = J$.

4 The Trilemma

We now turn to our main result, which is that one cannot jointly maintain that (i) prices satisfy no arbitrage, (ii) investors have preferences over payoffs, and (iii) supply shocks can be used to move along a demand curve without shifting it. Since Sections 2 and 3 already established the importance of the first two conditions, here we focus on showing that supply shocks generically fail to produce the price variation necessary to move along a stable demand curve.

We begin by defining the theoretical ideal of an experiment which allows the econometrician to identify the slope of an asset-level demand curve. We then compare this ideal experiment to the price variation by an exogenous supply shock, and show that the two are generally misaligned.

4.1 Ideal experiment

As shown in Section 2, canonical portfolio choice exhibits demand complementarities whereby the demand curve for any asset depends on the entire vector of asset prices. Measuring an asset-level demand elasticity thus requires an ideal experiment in which the investor faces *ceteris paribus* variation in a single asset price.

It is useful to describe the ideal experiment in terms of state prices, as these ultimately determine optimal consumption plans through the cost of consumption. The investor observes asset prices p and payoff matrix Y. Equation (2) allows the investor to infer the vector of state prices implied by prevailing asset prices:

$$q = Y^{+}p, \tag{3}$$

where Y^{+} is the Moore-Penrose pseudo-inverse of Y. If Y is square, as when markets are complete, then $Y^{+} = Y^{-1}$ and there is a unique vector of state prices. If markets are incomplete ($J < Z$), then there are many feasible state price vectors. We focus on the minimum norm solution with pseudo-inverse $Y^{+} = Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}$.

The ideal experiment consists of a pure price shock to a single asset. Equation (3) shows that, under no arbitrage, such a shock asset j implies a specific change to state prices which is fully determined by the inverse payoff matrix.

Lemma 2 (State price changes in the ideal experiment) *Let v_j denote the unit vector in $\mathbb{R}^J$ with 1 in the j-th position and zeros elsewhere. Then the changes in state prices* *given the exogenous variation in a single price p_j are*

$$\Delta \mathbf{q}_j^{\text{ideal}} \equiv \frac{\partial q}{\partial p_j} = Y^{+} v_j.$$

Proof. The assertion follows immediately from equation (3). ■

Identifying asset demand thus requires shocks which generate the state price variation $\Delta \mathbf{q}_j^{\text{ideal}}$ associated with the ideal experiment.

4.2 Measurement using supply shocks

In practice, one rarely observes direct shocks to prices themselves. Instead, one may observe shocks to an economic environment that trigger equilibrium price changes. As such, empirical approaches to estimating asset demand elasticities typically rely on suitably exogenous variation in the (residual) asset supply curve faced by a given investor. We will argue that this approach generally fails to generate the appropriate identifying variation in the context of asset markets.

To do so, we must describe how supply shocks affect state prices in a general class of models. Given the standard assumption of risk-averse preferences with decreasing marginal utility, we study settings in which a positive supply shock to asset *j* must reduce state prices in all states where asset *j* has a strictly positive payoff. We say that these settings exhibit downward-sloping consumption demand.

Definition 2 (Downward-sloping consumption demand) *Let $E \equiv (E_j)_{j=1}^{J} \in \mathbb{R}_{++}^{J}$ denote the vector of aggregate asset endowments. An economy has downward-sloping consumption demand if there exists a $Z \times Z$ matrix V with strictly positive diagonal elements such that*

$$\Delta \mathbf{q}_j^{\text{supply}} \equiv \frac{\partial q}{\partial E_j} = -V y_j^{\mathrm{T}} \qquad \text{for all assets } j,$$

where y_j^{T} is the transpose of the j-th row $y_j \equiv (y_j(z))_{z=1}^{Z}$ of Y.

In this definition, V captures the marginal change in the market-wide pricing kernel, which is taken as given by each individual investors. That V has strictly positive diagonal elements then captures our assumption that increases in the supply of state-contingent payoffs reduce the marginal price of these payoffs.

Definition 2 imposes no assumptions on V's off-diagonal entries. In economic terms, these entries capture potential *direct* preference-based spillovers across state prices in response to a supply shocks. Whether such spillovers exists depends on the economic model. The canonical model with additive separable utility over consumption in different states of the world (as in Section 2) has zero off-diagonal elements. Example 1 illustrates this with a representative investor. Non-separable models such as recursive utility (Epstein and Zin, 1989; Kreps and Porteus, 1978) or general aggregators instead generally imply non-zero off-diagonal elements.

As we have argued, the central identification challenge for demand estimation in asset markets is the threat of cross-asset spillovers. Therefore, the identification challenge is generically *weaker* when there are no direct, preference-based spillovers in state prices. To provide favorable conditions for identification, we thus assume that no such spillovers exist. This is Assumption 2.

Assumption 2 (No Direct Spillovers Across State Prices) *The marginal pricing kernel V is a diagonal matrix. Hence there are no direct state price spillovers.*

The only case in which non-diagonal V might help to identify demand is the knife-edge case where the preference-based spillovers in V just so happen to exactly offset the cross-asset restrictions implied by no arbitrage. However, V is determined by preferences and the aggregate supply of state-contingent payoffs while the no-arbitrage relation depends only on the payoff matrix Y. Hence there is no economic reason for such an offset to occur.[2]

Example 1 (V in an additive separable representative-agent model) *In a standard representative-agent model with additive separable preferences over consumption, state*

[2]Indeed, Section 5 uses random matrix theory to show that the sign of each element of Y^+ is close to a coin flip, with odds that depend only on the payoff matrix. Hence small perturbations to the payoff matrix can flip the sign of an element in Y^+ without much altering V.

prices relate to marginal utility over aggregate consumption,

$$\frac{\partial q_z}{\partial E_j} = \frac{\delta}{1-\delta}\pi_z \frac{u''(C_z)}{u'(C_0)} y_j(z) < 0,$$

where C_0 and C_z are aggregate consumption at date 0 and in state z. Thus the marginal price kernel is a strictly positive diagonal matrix,

$$V = -\frac{\delta}{1-\delta}\text{diag}\left(\pi_1 \frac{u''(C_1)}{u'(C_0)}, \ldots, \pi_z \frac{u''(C_z)}{u'(C_0)}, \ldots, \pi_Z \frac{u''(C_Z)}{u'(C_0)}\right).$$

4.3 Supply Shocks Do Not Generate the Ideal Experiment

We now show that supply shocks generically fail to produce the ideal experiment.

We study two definitions of alignment between supply shocks and the ideal experiment. The first is that the supply shock generates *exactly* the required variation in state prices, up to a scalar multiple to adjust the magnitude of the shock. This condition is necessary to ensure that supply shocks permit exact identification of demand functions for financial assets.

Condition 1 (Identical variation) *A supply shock to asset j generates the ideal state price variation for asset j if there exists some scalar k_j such that*

$$\Delta \mathbf{q}_j^{\text{ideal}} = k_j \Delta \mathbf{q}_j^{\text{supply}}.$$

This condition holds for all assets if and only if

$$Y^+ = -VY^{\mathrm{T}}K, \quad \textit{where } K \equiv \text{diag}(k_1, \ldots, k_J).$$

Even if Condition 1 fails, a supply shock may still provide useful variation if it does not depart too much from the ideal experiment. Hence we also consider a much weaker condition, namely the state price variation generated by a supply shock has the same *sign* as the state price changes in the ideal experiment.

Condition 2 (Variation of the same sign) *The supply shock generates state price variation of the same sign if $\Delta \mathbf{q}_j^{\text{ideal}}$ has the same sign as $\Delta \mathbf{q}_j^{\text{supply}}$ element by element. Given* *that Y has only weakly positive entries, this condition holds for all assets if Y^+ has only weakly positive entries.*

This condition is important because it ensures that the supply shock correctly induces the same *directional* pattern in state prices. If it fails, there are state-contingent payoffs which should become more expensive in the ideal experiment but actually become cheaper upon a supply shock. Such errors can naturally lead to large biases when estimating substitution.

We can then state our main result, which states that Conditions 1 and 2 are satisfied only under highly restrictive, non-generic conditions on the payoff matrix. In particular, for every state of the world there must exist a *unique* asset which offers a positive payoff in the world. Strikingly, both conditions require the *same stringent restrictions.* That is, as long as one wants to be sure to satisfy the minimal requirement that the induced state price variation is of the same sign as in the ideal experiment, then there must be no assets with overlapping payoffs.

Definition 3 (Overlapping payoffs) *Assets j and j' have overlapping payoffs if there exists at least one state of the world z such that $y_j(z) > 0$ and $y_{j'}(z) > 0$.*

Theorem 1 (Trilemma) *If Conditions 1 or 2 are satisfied, then YY^{T} is diagonal, and:*

(i) If YY^{T} is diagonal, then there are no assets with overlapping payoffs.

(ii) If markets are complete, then YY^{T} is diagonal if and only if Y is diagonal up to permutations.

Theorem 1 shows that there must be misalignment in magnitude and sign between the supply shock and the ideal experiment for at least one asset (i.e., one row of the payoff matrix). The next proposition strengthens this result by showing that such errors are guaranteed to occur for every asset.

Proposition 2 *If each column of Y has at least two strictly positive elements, then each column of the Moore-Penrose inverse Y^+ contains at least one negative element: for each $j \in \{1, \ldots, J\}$, there exists at least one $z \in \{1, \ldots, Z\}$ such that $(Y^+)_{z,j} < 0$.*

These conditions in Theorem 1 are unrealistic for almost all standard financial assets, as they require that there are no states of the world in which any given asset has positive payoffs while another asset also has positive payoffs. This is plainly violated for generic payoff distributions. It is therefore striking that, outside of these knife-edge restrictions, supply shocks do not even guarantee *directional* alignment with the ideal experiment. In the next section, we further document that directional errors are a pervasive problem for realistic payoff processes.

5 How Severe are these Problems?

The previous section established the generic misalignment between supply shocks and the ideal experiment without imposing any structure on the payoff matrix.

We now characterize the severity of this misalignment for realistic payoff processes in two steps. First, we analyze the asymptotic properties of factor-structured payoff processes. We find that approximately *half* of all entries in the inverse payoff matrix have the wrong sign, and that chance that any given individual entry is of the wrong sign is a coin flip. That is, small perturbations to the payoff matrix can alter the sign of the state price changes induced by a supply shock. Second, we conduct a simple empirical exercise using payoff data from S&P 500 stocks and show that it closely aligns with the theoretical findings.

5.1 Factor-structured Payoff Processes

We begin by studying theoretical properties of factor-structured payoff processes. As we will see, the presence of a factor structure is not chosen to create sign reversals, and may actually serve to reduce their prevalence. Even so, we find that, in the limit of many states and many assets, the share of sign reversals is well-approximated by *one half*. Simulations show that this remains a good approximation under a wide range of parameter configurations even when Z and J are not large. Hence sign reversals are a severe problem for realistic payoff processes.

Problem Statement. Because true payoffs are latent, we study random draws of Y generated from a factor structure. This allows us to characterize, in probability, the expected sign structure of its pseudo-inverse. Specifically, let payoff matrix $Y \in \mathbb{R}^{J \times Z}$ with $J \leq Z$ be defined by the following single factor structure, where $y_{j,z}$ represents the payoff of asset j in state z:

$$y_{j,z} = \alpha_j + \beta_j f_z + \varepsilon_{j,z} = \underbrace{\alpha_j + \beta_j \overline{f}}_{\equiv \gamma_j} + \beta_j (f_z - \overline{f}) + \varepsilon_{j,z}, \quad \text{where} \quad \overline{f} \equiv \mathbb{E}[f_z].$$

The Appendix shows that the analysis can be extended to multi-factor processes. We assume that the following conditions hold:

(A1) $(\alpha_j, \beta_j)_j$ are i.i.d., independent of $(f_z)_z$ and $(\varepsilon_{j,z})_{j,z}$, with finite second moments.

(A2) $(f_z - \overline{f})_z$ are i.i.d., with bounded, continuous, and symmetric densities around 0, with $\sigma_f^2 \equiv \mathbb{V}[f_z] < \infty$.

(A3) The idiosyncratic errors $(\varepsilon_{j,z})_{j,z}$ are i.i.d. across (j, z), with bounded, continuous, and symmetric densities around 0, with $\sigma_\varepsilon^2 \equiv \mathbb{V}[\varepsilon_{j,z}] > 0$. Also, factors and errors are mutually independent.

While we are interested in non-negative payoff matrices, $y_{j,z} \geq 0$, we do not explicitly impose extra assumptions on $(\alpha_j, \beta_j)_{j \geq 1}$, $(f_z - \overline{f})_z$, and $(\varepsilon_{j,z})_{j,z}$ to force $y_{j,z} \geq 0$. This is because Theorem 2 does not depend on this condition, and because truncating distributions to satisfy this restriction does not alter our result.

Let Y^+ denote the Moore-Penrose pseudo-inverse of Y.[3] Given that Y is random, we characterize the fraction of positive entries in Y^+ for given (J, Z),

$$p(J, Z) \equiv \frac{1}{JZ} \sum_{j=1}^{J} \sum_{z=1}^{Z} \mathbb{I}\left((Y^+)_{z,j} > 0\right). \tag{4}$$

[3] The rank of Y is J almost surely as long as the noise terms $(\varepsilon_{j,z})$ are drawn from a continuous distribution (which we assume). This is because the set of $J \times Z$ matrices with $\text{rank}(Y) < J (\leq Z)$ is of measure zero. Thus, $Y^+ = Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}$ almost surely.

If Y is weakly positive, $p(J)$ measures the fraction of state-price changes whose direction is correct under the ideal experiment. Hence it is a useful measure of the degree of alignment between supply shocks and ideal experiment.

The properties of small random matrices are difficult to characterize with any generality. Hence we study asymptotic properties of the share of positive entries as the number of states becomes large: $p(J) = \operatorname*{plim}_{Z\to\infty} p(J, Z)$. This aligns with empirical practice which often considers continuous payoff distributions. The existence of $p(J)$ is guaranteed by the law of large numbers. We then study its behavior as the number of assets grows large: $J \to \infty$. Later, we use simulations to show that our main results remain robust even away from these limits.

Result. We can then establish our main result in this section: the share of sign mismatches is well-approximated by *one half*. That is, for realistic payoff processes, directional errors are the norm, not an outlier.

Theorem 2 *Under Assumptions (A1)-(A3), there exists a constant C_1 such that, for almost every realization of $(\alpha_j, \beta_j)_j$, for sufficiently large J,*

$$p(J) = \frac{1}{2} + \frac{C_1}{J} + O\left(J^{-2}\right).$$

Consequently,

$$\operatorname*{plim}_{J\to\infty} p(J) = \frac{1}{2}.$$

Sketch of Proof. *We sketch the main argument. The full proof is in the Appendix. Theorem 2 states that the deviation of $p(J)$ from $\frac{1}{2}$ vanishes at a rate of $O(J^{-1})$, which we show depends on the factor loadings (α, β) and signal-to-noise ratio. The proof hinges on decoupling the asymptotic limits $Z \to \infty$ and $J \to \infty$. By letting the number of observations Z tend to infinity, the sample Gram matrix $\frac{1}{Z} YY^{\mathrm{T}}$ converges almost surely to the population second moment matrix Σ (which is invertible). Thus, the pseudo-inverse behaves as $Y^+ \approx \frac{1}{Z} Y^{\mathrm{T}} \Sigma^{-1}$, when Z is sufficiently large. Consequently, the sign of $(Y^+)_{z,j}$ is determined by the sign of $(\Sigma^{-1} y_z)_j$, where $y_z \equiv (y_{j,z})_j$ is the z-th column of Y.*

The core of the argument is the decomposition of the sign-determining variable $(\Sigma^{-1}y_z)_j$ into two components: (i) a small deterministic mean shift $\mu \equiv (\mu_j)_j$ and (ii) a dominating symmetric stochastic fluctuation $(W_{z,j})_j$. Letting $\gamma \equiv (\gamma_j)_j$, $\beta \equiv (\beta_j)_j$, and $\varepsilon_z \equiv (\varepsilon_{j,z})_j$, the column vector y_z can be expressed as:

$$y_z = \gamma + (f_z - \bar{f})\beta + \varepsilon_z.$$

Operating Σ^{-1} from the left, we can write, for each j, as:

$$(\Sigma^{-1}y_z)_j = \underbrace{(\Sigma^{-1}\gamma)_j}_{\mu_j \text{ (deterministic)}} + \underbrace{(f_z - \bar{f})(\Sigma^{-1}\beta)_j + (\Sigma^{-1}\varepsilon_z)_j}_{W_{z,j} \text{ (stochastic)}}. \tag{5}$$

Since the fluctuation $(W_{z,j})_{z,j}$ is symmetric around zero, the probability of a positive sign is exactly $\frac{1}{2}$ if the mean shift μ is zero. The difficulty in the proof is to show that even when $\mu > 0$ the distortion it creates is small and decreasing in J.

The asymptotic constant C_1 can be computed explicitly as:

$$C_1 = f_W(0) \cdot \Theta_1,$$

where (i) f_W is the probability density function for the stochastic fluctuation $(W_{z,j})_{z,j}$ and (ii) a constant Θ_1 is the sum of $(\mu_j)_j = \Sigma^{-1}\gamma$ (both of which are computed when $J \to \infty$, to obtain tractable expressions):[4]

$$\Theta_1 \equiv \lim_{J\to\infty} \sum_{j=1}^{J} (\Sigma^{-1}\gamma)_j = \frac{\mathbb{E}[\gamma]\,\mathbb{E}[\beta^2] - \mathbb{E}[\beta]\,\mathbb{E}[\gamma\beta]}{\mathbb{E}[\gamma^2]\,\mathbb{E}[\beta^2] - \mathbb{E}[\gamma\beta]^2}. \tag{6}$$

This shows that for given (but large) J the adjustment factor C_1 is proportional on the density around 0 since the higher the density the more distortion will be introduced from $\mu > 0$. Since the probability of $\mu \leq 0$ is $\frac{1}{2}$ (given the symmetry of f_W), the probability $p(J)$ is approximated by $\frac{1}{2} + \frac{C_1}{J}$ when J is fixed but large. ■

Perhaps more disconcertingly, the proof shows that the sign of each element

[4]Note that Assumption (A3) implies that $(\Sigma^{-1}y_z)_j$ has a continuous density at zero, since it is the sum of the deterministic term μ_j and the stochastic term $W_{z,j}$, where $W_{z,j}$ has a continuous density (being a linear combination of continuous random variables).

of the inverse payoff matrix is also a coin flip. In particular, each entry of Y^+ behaves asymptotically like a scaled draw from the symmetric random variable $(\Sigma^{-1}y_z)_j$, whose distribution is centered and continuous. Thus, the positive and negative tails of $(\Sigma^{-1}y_z)_j$ are mirror images and of equal magnitudes. This severe lack of stability makes it difficult to appropriately control for directional errors.

Note also that the specific distributions chosen for α, β, f, and ε only affect the magnitude of the constant C_1, not the fundamental asymptotic behavior. Similarly, in a setting with K factors, as long as properly extended versions of (A1)-(A3) hold, all that changes is that there will be a more complex constant C_K.

Calibration and Numerical Exploration To illustrate the $O(J^{-1})$ convergence rate and estimate the proportionality constant C_1, we performed Monte Carlo simulations using parameters that generate a share of idiosyncratic risk roughly consistent with the empirical data. Concretely, we assume:[5]

$$\alpha_j \sim \mathcal{U}[10,20], \qquad f_z \sim \mathcal{N}(1,\sigma_f^2) \quad \text{with} \quad \sigma_f = \frac{1}{2},$$
$$\beta_j \sim \mathcal{U}[0.5,1.5], \qquad \varepsilon_{j,z} \sim \mathcal{N}(0,\sigma_\varepsilon^2) \quad \text{with} \quad \sigma_\epsilon = 1.$$

As shown in Lemma 4, under Gaussian $(f_z, \varepsilon_{j,z})$, $f_W(0) = \frac{\sigma_\varepsilon}{\sqrt{2\pi}}$. From Equation (6) we have $\Theta_1 = 0.045$. Thus, $C_1 \approx 0.01795$. Hence even if our choice of σ_ε were off by an order of magnitude, the result would be practically the same.[6]

Figure 1, shows that the theoretical prediction for $Z \to \infty$ and large J can perform remarkably well even for moderate values of Z and small J.[7] Almost, half the elements of Y^+ have the wrong sign.

[5]The high values of $\alpha_j (\sim \mathcal{U}[10,20])$ effectively guarantee that all entries of Y are positive. Note, however, that our theoretical results do not require that. Also, truncation of the normal distributions for f and ε (to force Y to be always non-negative) do not qualitatively alter our results.

[6]Details and an interactive version of the code are available online.

[7]We took $Z = 1000$ as we vary the number of assets $J \in \{5, 10, \ldots, 500\}$. For the empirical frequency $p(J,Z)$, we took the average of 1000 runs (of the Monte Carlo simulations). While almost invisible, Figure 1 also depicts the 95% confidence interval. .

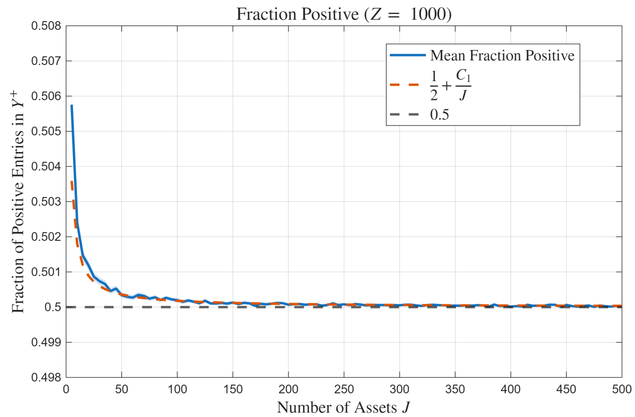

Figure 1: The Monte Carlo Simulation for the Sign Frequency of Y^+. The figure depicts the empirical frequency of positive entries $p(J, Z)$ (defined by (4)) and the theoretical approximation $\frac{1}{2} + \frac{0.01795}{J}$ when $Z = 1000$ for various J.

5.2 Empirical Validation

To further gauge the empirical relevance of our arguments, we conduct a simple excercise using stocks in the $S\&P$ 500. This exercise is not intended to be exhaustive, but rather a simple consistency check between our theory and the data.

Since the true payoff matrix is latent, we construct (subsets) of it by sampling realized payoffs. The sample consists of 428 stocks that remained in the $S\&P$ 500 from 2020 to 2024. The payoff for each stock is computed as the end-of-quarter price plus the sum of dividends paid during that quarter. We construct a 20×20 payoff matrix Y by randomly selecting 20 stocks (J). The columns (Z) correspond to the 20 quarterly payoff observations from 2020Q1 to 2024Q4. This yields a 20×20 payoff matrix with weakly positive entries. We then invert this payoff matrix and compute the share of negative entries in Y^+ as well as the relative magnitude of the negative and positive entries (in terms of the median and the maximum). We then repeat this exercise ten times with replacement.

Table 1 shows that our theoretical predictions hold remarkably well: the share of positive entries of Y^+ is approximately one half, and the negative entries are of equal magnitude. This again shows that the barriers to identification we document are generic and pervasive.

Metric (averaged over 10 iterations)	Value
Percentage of positive entries in Y^+	50.58%
Ratio: (abs negative-entry median) / (positive-entry median)	1.030
Ratio: (absolute negative minimum) / (positive maximum)	1.078

Table 1: Results of our empirical exercise averaged over 10 iterations.

6 Illustration in a General Equilibrium Model

The previous sections have established the generic and pervasive mismatch between the ideal experiment and supply shocks. We now illustrate the implications of this mismatch for errors in asset-level demand elasticity estimates. Since this requires a fully specified model, we study a simple example economy with a log-utility representative investor based on Fuchs, Fukuda, and Neuhann (2025a).[8]

Setup. Markets are complete. There are two assets and two states of the world, both denoted by g (green) and r (red). The probability of state $z \in \{g, r\}$ is $\pi_z \in (0, 1)$. The payoff profile of asset $j \in \{g, r\}$ is $y_j = (y_j(g), y_j(r))$. The aggregate endowments are given by $(e_0, e_g, e_r) = (1, 1 + s_g, 1)$, where s_g is a supply shock to the green asset. Table 2 depicts the payoff matrix.

Parameter $\epsilon \in (0, 1)$ determines the degree of complementarity between green and red assets. In the limit $\epsilon \to 0$, green and red assets are perfect substitutes with respect to their cash flows. The assets become more complementary as ϵ increases. In the limit $\epsilon \to 1$, the green and red assets are Arrow securities paying exactly one unit in one state of the world.

	State g (π_g)	State r (π_r)
Asset g	$\frac{1}{2}(1+\epsilon)$	$\frac{1}{2}(1-\epsilon)$
Asset r	$\frac{1}{2}(1-\epsilon)$	$\frac{1}{2}(1+\epsilon)$

Table 2: Payoff matrix.

[8]Log utility is convenient to obtain simpler expressions but not key to any of the results.

State Prices and Demand. Since markets are complete, we can solve the decision problem in terms of state-contingent consumption. Let c_z denote quantities of Arrow securities and q_z the associated state prices. The standard necessary and sufficient optimality condition for Arrow security $z \in \{g, r\}$ is

$$q_z = \pi_z \frac{\delta}{1-\delta} \frac{u'(c_z)}{u'(c_0)}. \tag{7}$$

Given the budget constraint, this condition determines optimal consumption as a function of Arrow prices. Let $W \equiv \left(2 + q_r(2 + (1-\epsilon)s_g) + q_g(2 + (1+\epsilon)s_g)\right)$ denote the investor's total wealth. Under log utility, optimal consumption is

$$c_0 = \frac{1-\delta}{2} W; \quad c_g = \frac{\delta \pi_g}{2 q_g} W; \quad c_r = \frac{\delta \pi_r}{2 q_r} W.$$

These optimal policies then uniquely determine the optimal asset positions.

Ideal experiment. Consider the ideal experiment where $s_g = 0$ and the investor faces an exogenous increase in the price of the green asset p_g while p_r remains fixed. Consistently with Lemma 2, the induced change in state prices is

$$\Delta \mathbf{q}_g^{\text{ideal}} = \frac{\partial}{\partial p_g} \begin{bmatrix} q_g \\ q_r \end{bmatrix} = \frac{1}{2\epsilon} \begin{bmatrix} 1+\epsilon \\ -(1-\epsilon) \end{bmatrix}. \tag{8}$$

A pure shock to p_g thus *raises* the cost of consumption in state g, but *lowers* it in state r. This decrease in q_r is necessary to keep p_r unchanged. Estimating the demand elasticity associated with this experiment requires a shock that triggers precisely this state price variation.

Supply shock. We now solve for the equilibrium prices after a supply shock and show that they do align with the ideal experiment. Market clearing requires consumption to equal available resources in every state:

$$c_z = y_g(z)(1 + s_g) + y_r(z).$$

Hence equilibrium state prices as a function of supply shock s_g are:

$$q_g = \pi_g \frac{\delta}{1-\delta} \cdot \frac{1}{1+\frac{1+\epsilon}{2}s_g} \qquad \text{and} \qquad q_r = \pi_r \frac{\delta}{1-\delta} \cdot \frac{1}{1+\frac{1-\epsilon}{2}s_g}. \tag{9}$$

In contrast to the ideal experiment, a negative supply shock to the green asset increases *both* state prices whenever $\epsilon < 1$. The reason is that the green asset pays off in both states of the world, so that the supply shock increases state-contingent consumption in both states. As such, the supply shock generates a state price change Δq_r that is of the *wrong sign* compared to the ideal experiment. The only exception is when both assets are Arrow securities ($\epsilon = 1$). This is the only case when the supply shock does not generate cross-asset spillovers. In particular, when $\epsilon < 1$, even a clean shock to one asset will trigger concurrent changes in the price of the other asset. Next, we show that this leads to a bias in asset-level demand elasticity that is greater when the assets are more substitutable (i.e., when ϵ is small).

Asset demand and implications for elasticity estimates. As we detail in Appendix B.4, the demand function for the green asset is

$$a_g(p_g, p_r) = \delta \frac{(1 + p_g(1+s_g) + p_r)\left((1-\epsilon^2)p_g - ((1+\epsilon)^2 - 4\epsilon\rho)p_r\right)}{(p_g - p_r)^2 - (p_g + p_r)^2\epsilon^2}. \tag{10}$$

In the ideal experiment, $s_g = 0$ and we observe a pure price shock to p_g. This gives us the standard own price elasticity formula:

$$\mathcal{E}_g^{\text{ideal}} \equiv -\frac{\partial a_g(p_g, p_r)}{\partial p_g} \frac{p_g}{a_g}.$$

When the price change is instead due to an infinitesimal supply shock, the resulting "elasticity" measure $\mathcal{E}_g^{\text{supply}}$ has an additional term which accounts for the effect of s_g on p_r:

$$\mathcal{E}_g^{\text{supply}} \equiv -\frac{\frac{da_g}{ds_g}}{\frac{dp_g}{ds_g}} \frac{p_g}{a_g} = \Big(-\frac{\partial a_g}{\partial p_g} - \frac{\partial a_g}{\partial p_r}\frac{\frac{dp_r}{ds_g}}{\frac{dp_g}{ds_g}}\Big)\frac{p_g}{a_g}.$$

Substituting for the equilibrium prices, these two measures are equal to:

$$\mathcal{E}_g^{\text{ideal}} = (1 + (1 - 2\pi_r)\epsilon) \frac{(1-\epsilon)^2 + 4\epsilon\pi_r(1-\delta\epsilon) + 4\delta\epsilon^2\pi_r^2}{8\pi_r(1-\pi_r)\epsilon^2};$$
$$\mathcal{E}_g^{\text{supply}} = (1 + (1 - 2\pi_r)\epsilon) \frac{2 - \delta(1 + (1 - 2\pi_r)\epsilon)}{(1+\epsilon)^2 - 4\epsilon\pi_r}.$$

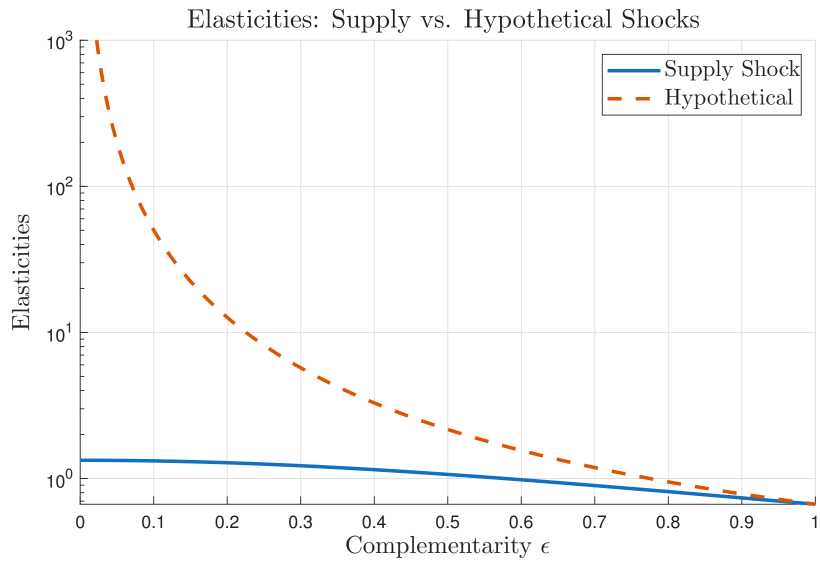

Figure 2: Ideal vs. supply-shock elasticities as a function of ϵ for $\delta = 2/3$ and $\pi_r = \pi_g = 1/2$. The ideal elasticity (solid line) diverges as $\epsilon \to 0$, while the supply-shock elasticity (dashed line) remains bounded. Both elasticities converge to $1 - \delta\pi_g = 2/3$ at the Arrow security limit $\varepsilon = 1$.

We plot both measures in Figure 2. The two differ by order of magnitude for small ϵ. In this range, the two assets are close substitutes. In the ideal experiment without price spillovers, this leads to very high demand elasticities with respect to a pure price shock. In the case of a supply shock, however, this very substitutability creates strong price spillovers that deter quantity changes on the equilibrium path. Hence, $\mathcal{E}_g^{\text{ideal}}$ diverges to infinity as $\epsilon \to 0$ while $\mathcal{E}_g^{\text{supply}}$ remains small. In contrast, when $\epsilon \to 1$ and the assets approach Arrow securities the measures converge. Thus, using supply shocks to estimate elasticities without properly accounting for the spillover effects leads to systemic underestimation of the true elasticity. The bias can be particularly significant in the presence of close substitutes. This can rationalize the findings of low "elasticities" in the literature.

There is no a priori reason to expect standard assets to correspond to the case of high ϵ. For example, our analysis in Section 5 suggests that positive and negative entries in the inverse payoff matrix are of roughly equal magnitude. In the context of our simple model here, this would indicate that ϵ is relatively small.

7 Addressing the Trilemma

Our results thus far establish a fundamental disconnect between the state price variation in the ideal experiment and the state price variation generated by asset-level supply shocks. We now evaluate whether and how researchers might potentially overcome these challenges using richer data or structural assumptions.

7.1 Multiple independent experiments

We begin by analyzing whether asset-level demand functions can be identified in an idealized setting where the researcher has access to multiple, *independent* quasi-experimental shocks to asset prices. To stack the deck in favor of identification, we assume that demand functions are approximately linear, so that demand functions can be described using a $J \times J$ substitution matrix. However, we caution that this is a very strong restriction when close substitutes are available.

A first-order approximation of investor i's demand system around $\overline{p}$ yields

$$a_i = \overline{a}_i + S_i(p - \overline{p}) + \varepsilon_i,$$

where $a_i \in \mathbb{R}^J$ is the vector of portfolio holdings, $p \in \mathbb{R}^J$ is the price vector, and ε_i is the vector of residual demand shocks. The asset-level *substitution matrix* $S_i \in \mathbb{R}^{J \times J}$ is the object of interest. Row k of matrix S_i collects the loadings of the demand for asset k on all prices, $(S_i)_k = \left(\frac{\partial a_{i,k}}{\partial p_1}, \ldots, \frac{\partial a_{i,k}}{\partial p_J}\right)$, while column j captures the derivative $\partial a_i / \partial p_j$. Under the assumption of linear demand, this derivative determines how the vector of asset quantities responds to changes in price p_j.

Suppose that the researcher has access to N distinct "experiments" indexed by n. Each experiment consists of a purely exogenous shock to the supply of a given asset (or combination of assets) which creates exogenous price changes. This generates a matrix of observable price changes G and that of quantity changes ΔA_i for each investor, defined as:

$$G \equiv \left[\Delta p^{(1)}, \ldots, \Delta p^{(N)} \right] \in \mathbb{R}^{J \times N};$$
$$\Delta A_i \equiv \left[\Delta a_i^{(1)}, \ldots, \Delta a_i^{(N)} \right] \in \mathbb{R}^{J \times N}.$$

Stacking the data from all N quasi-experiments yields the matrix equation

$$\Delta A_i \;=\; S_i \, G \;+\; U_i, \tag{11}$$

relating the observed quantity changes to the observed price changes and matrix of residual demand shocks $U_i \in \mathbb{R}^{J \times N}$. Our assumptions imply that $\mathbb{E}[U_i \mid G] = 0$.

We begin by establishing a positive identification result. In the theoretical ideal where the number of independent experiments equals the dimensionality of the asset span, ordinary least squares identifies the investors' substitution matrix.

Proposition 3 (Complete identification with J experiments) *Let the number of independent experiments equal the dimensionality of the asset span, so that the matrix of observed price changes is full row rank,* $\text{rank}(G) = J$*. Let $G^{+} (= G^{-1})$ denote the Moore-Penrose pseudo-inverse of G. Then the unique ordinary least-squares estimator of S_i is*

$$\hat{S}_i \;=\; \Delta A_i \, G^{\mathrm{T}} \left(G \, G^{\mathrm{T}} \right)^{-1} \;=\; \Delta A_i \, G^{+}, \tag{12}$$

where $\hat{S}_i$ is an unbiased and consistent estimator of S_i. When $U_i = 0$, $\hat{S}_i = S_i$.

Proposition 3 provides a constructive benchmark: with as many independent shocks as the dimensionality of the asset span, demand functions are point-identified under the (strong) assumption of linear demand. We refer to this result as *complete identification* because every element of S_i is point-identified. This provides one constructive method for asset demand estimation, which is to find settings with sufficiently many shocks relative to the number of assets.

However, these data requirements are stringent. In many applications, researchers observe far fewer than J independent experiments. For example, Koijen and Yogo (2019) rely on a single cross-section of prices and quantities, which corresponds to a single independent experiment. Increasing the number of experiment requires the existence of strongly segmented markets subject to market-specific shocks, or time series variation under stationarity. But most financial markets are not strictly segmented, and the market-wide pricing kernel typically contains a permanent component (Alvarez and Jermann, 2005; Borovička, Hansen, and Scheinkman, 2016). This lack of stationarity limits the scope of the time-series methods proposed in, e.g., Haddad, He, Huebner, Kondor, and Loualiche (2025). Another alternative is to combine structural models with data on bid *schedules*, not just equilibrium holdings and prices (Allen, Kastl, and Wittwer, 2025).

Given these limitations, we must assess the identification of substitution matrix S_i in the empirically relevant case where $N < J$. The next result shows that the substitution matrix is not point-identified if $N < J$, and indeed that demand parameters are arbitrarily unconstrained beyond the span of observed shocks.

Proposition 4 (Incomplete identification with $N < J$ experiments) *Let $P_G \equiv GG^+$ be the orthogonal projector onto $\text{col}(G)$, the column space of the matrix of observed price changes G, where $G^+ \equiv (G^{\mathrm{T}}G)^{-1}G^{\mathrm{T}}$. Then the general solution to the least-squares problem is*

$$S_i = \Delta A_i G^+ + B_i(I_J - P_G),$$

where $B_i \in \mathbb{R}^{J \times J}$ is an arbitrary matrix that is entirely unrestricted by the data and I_J is the identity matrix.

That is, any component of S_i in the null space of G is not point-identified and cannot be bounded without ex-ante theoretical restrictions which cannot be rejected by the data. What *is* identified is the *projection* of S_i onto observed shocks,

$$S_i P_G = \Delta A_i G^+ P_G = \Delta A_i G^+.$$

Our results show this projection does not identify the structural slope of any asset-level demand function because it is contaminated by correlated price changes.

Remark 3 (Relation to Collinearity and Weak Instruments) *When asset prices satisfy $p = Yq$, equilibrium price movements are confined to the low-dimensional space spanned by the state prices q. As a result, the matrix of observed price changes G is typically of rank deficient, implying that instruments constructed from asset-level supply shocks are highly collinear. In instrumental-variable terms, the first-stage regression of individual prices on such instruments is likely weak once other prices are controlled for: the conditional F-statistic is small even if unconditional correlations are large. However, the absence of weak instruments—that is, a strong first stage—is not sufficient for credible identification. Even when instruments generate large first-stage variation, they may still induce the wrong direction of price movements relative to the ideal experiment that isolates an own-price effect. In the terminology of Proposition 4, such instruments span an incorrect subspace of the price space, identifying only projections of demand elasticities rather than structural slopes. Hence, strong instruments ensure relevance but not alignment: they are necessary, but not sufficient, for consistent identification of asset-level demand.*

7.2 Portfolio Aggregation and Alternative Estimands

Which objects of interest can be identified when the identification of asset-level demand curves is infeasible?

One approach is to estimate elasticities over *portfolios* rather than individual assets. While our formal results apply equally to any asset or portfolio, there are some potential benefits and costs of such aggregation. The main benefit is a reduction in the dimensionality of the choice set. Following Section 7.1, this means that one needs fewer independent shocks to identify a given substitution matrix. A disadvantage is that aggregation into large portfolios (e.g., stocks and bonds) may make it more difficult to find supply shocks that are suitably exogenous to demand. Moreover, if shocks occur at the asset level, one must construct a portfolio-level shock by combining different asset-level shocks. The appropriate weighting depends on payoffs, and thus requires knowledge of the latent payoff matrix. For example, Binsbergen, David, and Opp (2025) use a structural model to reverse-engineer the set of shocks needed to generate a particular price change.

Beyond simple aggregation, one might combine assets into portfolios which resemble Arrow securities. As before, the main challenge is that this requires knowledge of the latent payoff matrix. An (2025) and An and Huber (2024) pursue a related approach by constructing portfolios orthogonalized with respect to returns and flows to specific investors. However, orthogonal payoffs are not sufficient to ensure no overlap in the payoff distribution. Their approach thus requires additional assumptions to eliminate spillovers, such as asserting that uncorrelated portfolios exhibit no spillovers and that the risk-free rate is exogenously fixed. Consistent with our results, this is an a-priori restriction on substitution patterns.

Lastly, one may be content to identify objects other than the asset-level demand elasticity. For example, Haddad, He, Huebner, Kondor, and Loualiche (2025) propose a specific conditional homogeneity restrictions on the (endogenous) substitution matrix. Under this assumption, they show that asset supply shocks can identify a "relative elasticity"—the difference between an asset's own- and cross-price elasticities relative to similar assets—but not the absolute elasticity. This circumvents the problem of cross-asset spillovers by estimating a different economic object. Nevertheless, identification is contingent on a-priori assumptions on unobservables, and small misspecification can lead to large biases.[9]

7.3 Structural Assumptions

Our results show that supply shocks generically fail to produce the price variation required to non-parametrically estimate asset demand functions. This suggests an important role for structural models in asset demand estimation. These models must be designed to account for the cross-asset interactions which underlie asset pricing and portfolio choice. For example, Fuchs, Fukuda, and Neuhann (2025a) shows that misspecification of substitution patterns can lead to large and systematic biases in the logit asset demand model proposed by Koijen and Yogo (2019). Moreover, because (arbitrarily) many theoretical models may be consistent with

[9] See Fuchs, Fukuda, and Neuhann (2025b) for a discussion on robustness to small deviations when assets are highly substitutable.

the observed data, these models must be judged on ex-ante theoretical considerations and plausibility, not on their empirical fit.

8 Conclusion

We provide a general analysis of the scope for demand estimation in asset markets. Our main conclusion is that asset demand analysis is sharply constrained by two foundational principles of asset pricing: investors ultimately care about asset payoffs; and asset prices should admit no arbitrage. These results are independent of specific assumptions on preferences, payoffs, and the economic environment. Our results highlight the importance of structural modeling in asset demand analysis, but also caution that such models must be carefully designed to account for the cross-asset interactions which lie at the heart of asset demand.

References

ALLEN, J., J. KASTL, AND M. WITTWER (2025): "Estimating Demand Systems with Bidding Data," Working Paper. [Cited on pages 6 and 29.]

ALVAREZ, F., AND U. J. JERMANN (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," Econometrica, 73(6), 1977–2016. [Cited on pages 4 and 29.]

AN, Y. (2025): "Demand Elasticity for Portfolios," Working paper, Johns Hopkins University. [Cited on pages 6 and 31.]

AN, Y., AND A. W. HUBER (2024): "Demand Propagation Through Traded Risk Factors," Discussion paper. [Cited on pages 6 and 31.]

BERRY, S. T., AND P. A. HAILE (2021): "Foundations of Demand Estimation," Handbook of Industrial Organization, 4(1), 1–62. [Cited on page 5.]

BINSBERGEN, J. H. V., B. DAVID, AND C. C. OPP (2025): "How (Not) to Identify Demand Elasticities in Dynamic Asset Markets," Working paper. [Cited on pages 6 and 30.]

BOROVIČKA, J., L. P. HANSEN, AND J. A. SCHEINKMAN (2016): "Misspecified Recovery," Journal of Finance, 71(6), 2493–2544. [Cited on pages 4 and 29.]

DUFFIE, D. (2001): Dynamic Asset Pricing Theory. Princeton University Press, third edn. [Cited on pages 10, 11, and 3.]

EPSTEIN, L. G., AND S. E. ZIN (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," Econometrica, 57(4), 937–969. [Cited on page 14.]

ERSHOV, D., J.-W. LALIBERTÉ, M. MARCOUX, AND S. ORR (2024): "Estimating Complementarity With Large Choice Sets: An Application to Mergers," Working Paper. [Cited on page 5.]

FOSGERAU, M., J. MONARDO, AND A. DE PALMA (2024): "The Inverse Product Differentiation Logit Model," American Economic Journal: Microeconomics, 16(4), 329–370. [Cited on page 5.]

FUCHS, W., S. FUKUDA, AND D. NEUHANN (2025a): "Demand-System Asset Pricing: Theoretical Foundations," Working paper, University of Texas at Austin and Bocconi University. [Cited on pages 6, 23, and 31.]

——— (2025b): "A Three-Assets Example with a Slight Perturbation," Working paper available here. [Cited on page 31.]

HADDAD, V., Z. HE, P. HUEBNER, P. KONDOR, AND E. LOUALICHE (2025): "Causal Inference for Asset Pricing," Discussion paper. [Cited on pages 6, 29, and 31.]

HARRIS, L., AND E. GUREL (1986): "Price and Volume Effects Associated with Changes in the S&P 500 List: New Evidence for the Existence of Price Pressures," Journal of Finance, 41(4), 815–829. [Cited on page 5.]

HE, Z., P. KONDOR, AND J. S. LI (2025): "Demand Elasticity in Dynamic Asset Pricing," Working paper. [Cited on page 6.]

IARIA, A., AND A. WANG (2020): "Identification and Estimation of Demand for Bundles," Working Paper. [Cited on page 5.]

KOIJEN, R. S. J., AND M. YOGO (2019): "A Demand System Approach to Asset Pricing," Journal of Political Economy, 127(4), 1475–1515. [Cited on pages 3, 6, 10, 28, and 31.]

——— (2025): "On the Theory and Econometrics of (Demand System) Asset Pricing," Working Paper. [Cited on page 9.]

KREPS, D. M., AND E. L. PORTEUS (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," Econometrica, 46(1), 185–200. [Cited on page 14.]

MAS-COLELL, A. (1977): "The Recoverability of Consumers' Preferences from Market Demand Behavior," Econometrica, 45(6), 1409–1430. [Cited on page 3.]

PLEMMONS, R. J., AND R. E. CLINE (1972): "The Generalized Inverse of a Nonnegative Matrix," Proceedings of the American Mathematical Society, 31(1), 46–50. [Cited on page 35.]

ROSS, S. A. (2004): Neoclassical Finance. Princeton University Press, Princeton, NJ. [Cited on page 10.]

SHLEIFER, A. (1986): "Do Demand Curves for Stocks Slope Down?," Journal of Finance, 41(3), 579–590. [Cited on page 5.]

STARKS, L. T. (2023): "Presidential Address: Sustainable Finance and ESG Issues - Value versus Values," Journal of Finance, 78(4), 1837–1872. [Cited on page 9.]

TOBIN, J. (1969): "A General Equilibrium Approach to Monetary Theory," Journal of Money, Credit and Banking, 1(1), 15–29. [Cited on page 4.]

WANG, A. (2024): "Demand with Complementarity: A Market-Share Approach," Working Paper. [Cited on page 5.]

A Appendix

A.1 Section 4.3

Proof of Theorem 1. First, we show that Condition 1 implies that YY^{T} is diagonal. Suppose $Y^{+} = -VY^{\mathrm{T}}K$ for some diagonal matrix $K \equiv \mathrm{diag}(k_1, \ldots, k_J)$. Operating Y on both sides from the left,

$$I_J = -YVY^{\mathrm{T}}K.$$

If $k_j = 0$ for some j, then the j-th column of K is the zero vector, and so is the j-th column of the right-hand side, which is impossible. Thus, $k_j \neq 0$ for all j. Then, YVY^{T} is a diagonal matrix:

$$\begin{cases} \sum_{z=1}^{Z} y_j(z) v_z y_{j'}(z) \neq 0 & \text{if } j = j' \\ \sum_{z=1}^{Z} y_j(z) v_z y_{j'}(z) = 0 & \text{if } j \neq j' \end{cases}.$$

Since $y_j(z), y_{j'}(z) \geq 0$, and $v_z > 0$, it follows that

$$\begin{cases} \sum_{z=1}^{Z} y_j(z) y_{j'}(z) \neq 0 & \text{if } j = j' \\ \sum_{z=1}^{Z} y_j(z) y_{j'}(z) = 0 & \text{if } j \neq j' \end{cases}.$$

Hence, YY^{T} is diagonal.

Second, we show that, more generally, Condition 2 implies that YY^{T} is diagonal. By Condition 2, the Moore-Penrose pseudo-inverse $Y^{+} = Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}$ is non-negative. By Plemmons and Cline (1972, Theorem 1), the pseudo-inverse Y^{+} is non-negative if and only if there exists a diagonal matrix with positive elements $D \equiv \mathrm{diag}(d_1, \ldots, d_Z)$ such that

$$Y^{+} = DY^{\mathrm{T}}. \tag{13}$$

Then, operating Y from the left,

$$I_J = YDY^{\mathrm{T}}.$$

Then, extracting the (j, k) element (with $j \neq k$) from each of both sides,

$$0 = \sum_{z=1}^{Z} y_j(z) d_z y_k(z).$$

Since $y_j(z) \geq 0$, $d_z > 0$, and $y_k(z) \geq 0$ for all $z \in \{1, \dots, Z\}$, it follows that

$$y_j(z) y_k(z) = 0 \text{ for all } z \in \{1, \dots, Z\}.$$

This implies that the (j, k) element (with $j \neq k$) of YY^{T} is 0:

$$0 = \sum_{z=1}^{Z} y_j(z) y_k(z). \tag{14}$$

Thus, YY^{T} is a diagonal matrix.

Third, we show that, given that YY^{T} is diagonal, there are no assets with overlapping payoffs. Since YY^{T} is invertible, it is a diagonal matrix with positive elements. Equation (14) implies that, for any $z \in \{1, \dots, Z\}$, there exists at most one $j \in \{1, \dots, J\}$ such that $y_j(z) > 0$.

Fourth, we show that if markets are complete then YY^{T} is diagonal if and only if Y has exactly one non-zero element in each row and in each column (so that Y is a diagonal matrix up a re-ordering of rows or columns). If YY^{T} is diagonal, then its (j, k) element is:

$$\begin{cases} \sum_{z=1}^{Z} y_j(z) y_j(z) > 0 & \text{if } j = k \\ \sum_{z=1}^{Z} y_j(z) y_k(z) = 0 & \text{if } j \neq k \end{cases}.$$

Hence, for each row j, there exists exactly one element z such that $y_j(z) > 0$. Thus, Y has J non-zero elements. Since Y is square and invertible, for each column z, there exists exactly one element j such that $y_j(z) > 0$.

Conversely, if Y has exactly one non-zero element in each row and in each column, then

$$\begin{cases} \sum_{z=1}^{Z} y_j(z) y_j(z) > 0 & \text{if } j = k \\ \sum_{z=1}^{Z} y_j(z) y_k(z) = 0 & \text{if } j \neq k \end{cases}.$$

Thus, YY^{T} is diagonal. ■

Remark 4 (Proof of Theorem 1) *Two remarks on the proof of Theorem 1 are in order. First, if YY^{T} is diagonal, then since YY^{T} is invertible under Assumption 1, $(YY^{\mathrm{T}})^{-1}$ is a diagonal matrix with positive entries. Since Y is non-negative, so is Y^{T}. Then, $Y^{+} = Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}$ is non-negative.*

Second, when each column of Y is not a zero vector, i.e., for each $z \in \{1, \dots, Z\}$, there exists at least one $j \in \{1, \dots, J\}$ such that $Y_{j,z} = y_j(z) > 0$, it can be shown that the diagonal matrix D in expression (13) is unique.

Proof of Proposition 2. Let y_j denote the j-th row of Y. Let y_k^{+} denote the k-th column of Y^{+}. It follows from $YY^{+} = I_J$ that:

$$\sum_{z=1}^{Z} y_k(z) Y^{+}_{z,k} = 1 \quad \text{for all} \quad k \in \{1, \dots, J\}; \tag{15}$$

$$\sum_{z=1}^{Z} y_j(z) Y^{+}_{z,k} = 0 \quad \text{if} \quad j \neq k. \tag{16}$$

Suppose to the contrary that there exists a column k in Y^{+} such that $y_k^{+} \geq 0$ element-by-element.

Consider the orthogonality condition (16) for some $j \neq k$. Since Y is non-negative, $y_k \geq 0$. We assumed $y_k^{+} = (Y^{+}_{z,k})_z \geq 0$. Thus, if $y_j(z) > 0$ then $Y^{+}_{z,k} = 0$. This must hold for all $j \neq k$. Therefore, y_k^{+} must be zero at any index z where any other row of Y is positive.

Now consider the normalization condition (15). For the sum to be strictly positive, there must exist at least one index z^* such that:

$$y_k(z^*) > 0 \quad \text{and} \quad Y^{+}_{z^*,k} > 0. \tag{17}$$

However, we know that $Y^{+}_{z^*,k} > 0$ is only possible if $y_j(z^*) = 0$ for all $j \neq k$. Combining this with expression (17), we see that index z^* represents a column in Y where: the entry in row k is positive: $y_k(z^*) > 0$; and the entries in all other rows i are zero: $y_i(z^*) = 0$ for $i \neq k$. This implies that column z^* of matrix Y has exactly one strictly positive element, which is a contradiction to the assumption of the statement. ■

A.2 Section 5: Proof of Theorem 2

Proof Overview. The proof consists of four steps. The first step establishes the asymptotic limit (i.e., the population covariance matrix) Σ of the Gram matrix $\frac{1}{Z}Y^{\mathrm{T}}Y$ as $Z \to \infty$. This allows the pseudo-inverse $Y^{+} = Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}$ to be approximated by $\frac{1}{Z}Y^{\mathrm{T}}\Sigma^{-1}$. The second step shows that each column of $Y^{\mathrm{T}}\Sigma^{-1}$ can be decomposed into the deterministic shift (i.e., $(\mu_j)_j$ in the main text) and the stochastic component centered around 0 (i.e., $(W_{z,j})_{z,j}$ in the main text). The third step establishes the sense in which the deterministic shift is small compared to the stochastic component $(W_{z,j})_{z,j}$ when J is large by applying the Woodbury identity to Σ. The fourth step computes the closed-form formula for the constant C_1 so that $p(J)$ is approximated by $\frac{1}{2} + \frac{C_1}{J}$ when J is large.

Step 1. In the first step, we replace the sample Gram matrix $G_Z \equiv \frac{1}{Z}YY^{\mathrm{T}}$ with the population covariance matrix Σ by the law of large numbers. Namely, as $Z \to \infty$ with J fixed, the sample covariance matrix G_Z converges almost surely to the population second moment matrix Σ (conditional on α and β), where

$$\Sigma = \gamma\gamma^{\mathrm{T}} + \sigma_f^2\beta\beta^{\mathrm{T}} + \sigma_\varepsilon^2 I_J = \sigma_\varepsilon^2 I_J + UU^{\mathrm{T}} \quad \text{with} \quad U \equiv \begin{bmatrix}\gamma & \sigma_f\beta\end{bmatrix} \in \mathbb{R}^{J\times 2}$$

is a rank-two perturbation of a scaled identity matrix. Note that Σ is positive definite so that it is invertible. This allows the pseudo-inverse Y^{+} to be approximated by $\frac{1}{Z}Y^{\mathrm{T}}\Sigma^{-1} = \frac{1}{Z}\Sigma^{-1}Y^{\mathrm{T}}$. Lemma 3 in the second step formally shows that the sign of $(Y^{+})_{z,j}$ is determined by the sign of the variable $(\Sigma^{-1}y_z)_j$. To that end, the second step starts by decomposing $(\Sigma^{-1}y_z)_j$ into the deterministic shift and the stochastic part symmetric around 0.

Step 2. Writing $y_z = \gamma + \beta(f_z - \bar{f}) + \varepsilon_z$ as in the main text, one can express

$$(\Sigma^{-1} y_z)_j = \underbrace{(\Sigma^{-1}\gamma)_j}_{\mu_j} + \underbrace{(f_z - \overline{f})(\Sigma^{-1}\beta)_j + (\Sigma^{-1}\varepsilon_z)_j}_{W_{z,j}}.$$

Conditional on the loadings (α, β), the term μ_j is a deterministic shift and the term $W_{z,j}$ is symmetric around zero.

Let $F_{W,j}$ be the CDF of $W_{z,j}$ conditional on loadings (α, β). Then,

$$\begin{aligned} \mathbb{P}((\Sigma^{-1} y_z)_j > 0) &= \mathbb{P}(\mu_j + W_{z,j} > 0) \\ &= 1 - F_{W,j}(-\mu_j) = \frac{1}{2} + f_{W,j}(0)\mu_j + O(\mu_j^2), \end{aligned} \tag{18}$$

where the last equality follows from the Taylor approximation of $1 - F_{W,j}(\cdot)$ and $F_{W,j}(0) = \frac{1}{2}$ (which follows because $W_{z,j}$ is symmetric around zero).

With these in mind, we now establish Lemma 3, which guarantees that the replacement of $Y^+ = \frac{1}{Z} Y^{\mathrm{T}} G_Z^{-1}$ with $\Sigma^{-1} Y^{\mathrm{T}}$ does not change the limiting sign frequency: since $G_Z^{-1} \to \Sigma^{-1}$ with $\|G_Z^{-1} - \Sigma^{-1}\| = O(Z^{-1/2})$, the difference between the two matrices vanishes in operator norm, and any potential sign disagreement occurs only when an entry of $(\Sigma^{-1} y_z)_j$ lies in a vanishing neighborhood of zero. Since replacing G_Z^{-1} by Σ^{-1} changes each entry by at most $O(Z^{-1/2})$, a sign disagreement can occur only with probability $o(1)$. This ensures that the asymptotic sign frequency is unaffected by the finite-Z approximation. Formally:

Lemma 3 (Population Replacement) *Fix J and (α, β). Then,*

$$\lim_{Z\to\infty} \max_{1 \le j \le J} \left| \frac{1}{Z} \sum_{z=1}^{Z} \mathbb{I}\Big(\big(\frac{1}{Z} y_z^{\mathrm{T}} G_Z^{-1}\big)_j > 0 \Big) - \mathbb{P}\Big(\big(\Sigma^{-1} y_z\big)_j > 0 \Big) \right| = 0 \quad a.s.$$

Consequently, conditional on (α, β), the probability $p(J)$ satisfies:

$$p(J) = \frac{1}{J} \sum_{j=1}^{J} \mathbb{P}\big((\Sigma^{-1} y_z)_j > 0\big). \tag{19}$$

Proof of Lemma 3. By the law of large numbers, $G_Z \to \Sigma$ a.s. Hence, G_Z is positive definite for large Z and $\|G_Z^{-1} - \Sigma^{-1}\| \to 0$ a.s. Let

$$D_{z,j} \equiv \left(\frac{1}{Z} y_z^{\mathrm{T}} G_Z^{-1}\right)_j - \left(\frac{1}{Z}\Sigma^{-1} y_z\right)_j = \frac{1}{Z} y_z^{\mathrm{T}} (G_Z^{-1} - \Sigma^{-1}) v_j,$$

where v_j is the unit vector in the j-th coordinate.

For any $\eta > 0$ and all large Z, $\|G_Z^{-1} - \Sigma^{-1}\|_F \leq \eta$ a.s., so $|D_{z,j}| \leq \frac{\eta}{Z}\|y_z\|$. A sign can flip only if $|(\frac{1}{Z}\Sigma^{-1} y_z^{\mathrm{T}})_j| \leq |D_{z,j}|$. Since $(\Sigma^{-1} y_z)_j = \mu_j + W_{z,j}$ has a continuous density at around 0 with value $f_{W,j}(0)$, we have:

$$\mathbb{P}\left(\left|(\Sigma^{-1} y_z)_j\right| \leq \delta\right) \leq 2 f_{W,j}(0)\delta + o(\delta) \quad (\delta \downarrow 0).$$

Since the inequality holds uniformly across $j \in \{1, \ldots, J\}$, the sign disagreement probability vanishes uniformly across the entire cross-section $j \in \{1, \ldots, J\}$. Taking $\delta = \frac{\eta}{Z}\|y_z\|$ and averaging over z (using $Z^{-1}\sum_z \|y_z\| \to \mathbb{E}\|y_z\|$ a.s.) shows the empirical fraction of sign disagreements is $O(\eta)$ a.s. Letting $\eta \downarrow 0$ proves the lemma. ■

So far, expressions (18) and (19) imply that

$$\begin{aligned} p(J) &= \frac{1}{J}\sum_{j=1}^{J}\left(\frac{1}{2} + f_{W,j}(0)\mu_j + O(\mu_j^2)\right) \\ &= \frac{1}{2} + f_{W,j}(0)\frac{1}{J}\sum_{j=1}^{J}\mu_j + \frac{1}{J}\sum_{j=1}^{J} O(\mu_j^2). \end{aligned}$$

As a preview, the third step shows that $\mu_j = O(J^{-1})$ and $f_{W,j}(0) \to f_W(0)$ uniformly so that one can write

$$p(J) = \frac{1}{2} + f_W(0)\frac{1}{J}\sum_{j=1}^{J}\mu_j + O(J^{-2}). \tag{20}$$

The fourth step find the closed-form expression for (6), i.e., $\Theta_1 \equiv \lim_{J\to\infty}\sum_{j=1}^{J}\mu_j$, to obtain

$$p(J) = \frac{1}{2} + f_W(0)\cdot\Theta_1\cdot\frac{1}{J} + O(J^{-2}).$$

Step 3. In the third step, we determine the magnitude of the mean-shift term $\mu_j = (\Sigma^{-1}\gamma)_j$. To that end, we apply the Woodbury matrix identity to obtain:

$$\Sigma^{-1} = (\sigma_\varepsilon^2 I_J + UU^{\mathrm{T}})^{-1} = \sigma_\varepsilon^{-2}\Big(I_J - U\big(I_2 + \sigma_\varepsilon^{-2}U^{\mathrm{T}}U\big)^{-1}\sigma_\varepsilon^{-2}U^{\mathrm{T}}\Big). \tag{21}$$

Since the 2×2 matrix $U^{\mathrm{T}}U$ satisfies

$$U^{\mathrm{T}}U = J\begin{bmatrix} \mathbb{E}_J[\gamma^2] & \sigma_f \mathbb{E}_J[\gamma\beta] \\ \sigma_f \mathbb{E}_J[\gamma\beta] & \sigma_f^2 \mathbb{E}_J[\beta^2] \end{bmatrix},$$

where $\mathbb{E}_J$ denotes the empirical mean over $j \in \{1, \dots, J\}$, each entry is of order $O(J)$. Thus, $\sigma_\varepsilon^{-2}U^{\mathrm{T}}U = O(J)$, which implies that the dominant term in the matrix $I_2 + \sigma_\varepsilon^{-2}U^{\mathrm{T}}U$ is the $O(J)$ contribution from $\sigma_\varepsilon^{-2}U^{\mathrm{T}}U$. Thus, when J is large, the 2×2 matrix $\big(I_2 + \sigma_\varepsilon^{-2}U^{\top}U\big)^{-1}$ is of order $O(J^{-1}) = O(1) \cdot O(J^{-1}) \cdot O(1)$. Consequently, each entry in the matrix $U\big(I_2 + \sigma_\varepsilon^{-2}U^{\top}U\big)^{-1}\sigma_\varepsilon^{-2}U^{\mathrm{T}}$ is of order $O(J^{-1})$, meaning that Σ^{-1} is asymptotically diagonal with off-diagonal entries that vanish at the same rate. Economically, the pseudo-inverse suppresses variation along the factor directions while leaving idiosyncratic risk largely unaffected. With this in mind, we establish:

Lemma 4 (Small Deterministic Shift) *The following hold uniformly in j:*

$$(\Sigma^{-1}\gamma)_j = O(J^{-1}), \qquad (\Sigma^{-1}\beta)_j = O(J^{-1}), \qquad (\Sigma^{-2})_{j,j} \to \sigma_\varepsilon^{-4}.$$

Consequently,

$$\sigma_{W,j}^2 = \sigma_f^2\big[(\Sigma^{-1}\beta)_j\big]^2 + \sigma_\varepsilon^2(\Sigma^{-2})_{j,j} \longrightarrow \sigma_\varepsilon^{-2} \quad \textit{uniformly in} \quad j.$$

If $(f_z, \varepsilon_{j,z})$ are Gaussian case, then

$$f_{W,j}(0) \longrightarrow \frac{\sigma_\varepsilon}{\sqrt{2\pi}} \quad \textit{uniformly in} \quad j.$$

Proof of Lemma 4. Since we have established $(I_2 + \sigma_\varepsilon^{-2}U^{\mathrm{T}}U)^{-1} = O(J^{-1})$, substituting this back into Woodbury identity (21) and noting that $\gamma = Uv_1$ yields

$$\Sigma^{-1}\gamma = \sigma_\varepsilon^{-2}\gamma - \sigma_\varepsilon^{-2}U\big(I_2 + \sigma_\varepsilon^{-2}U^{\mathrm{T}}U\big)^{-1}\sigma_\varepsilon^{-2}U^{\mathrm{T}}\gamma.$$

The first term $\sigma_\varepsilon^{-2}\gamma$ is $O(1)$ in each component. However, since γ lies in the column space of U, we have $U^{\mathrm{T}}\gamma = U^{\mathrm{T}}Uv_1$, which is $O(J)$. Thus, the second term equals

$$\sigma_\varepsilon^{-4}U\big(I_2 + \sigma_\varepsilon^{-2}U^{\mathrm{T}}U\big)^{-1}U^{\mathrm{T}}Uv_1 = \sigma_\varepsilon^{-2}U \cdot O(J^{-1}) \cdot O(J) = \sigma_\varepsilon^{-2}U \cdot O(1).$$

Each component of this correction term is $O(1)$, and it precisely cancels the leading $O(1)$ term $\sigma_\varepsilon^{-2}\gamma$. What remains is a residual of order $O(J^{-1})$: each component $\mu_j = (\Sigma^{-1}\gamma)_j$ satisfies $|\mu_j| = O(J^{-1})$ uniformly in j. The same reasoning applies to β, giving $(\Sigma^{-1}\beta)_j = O(J^{-1})$.

Next, squaring expression (21) gives

$$\Sigma^{-2} = \sigma_\varepsilon^{-4}\Big(I_J - 2UA_J^{-1}\sigma_\varepsilon^{-2}U^{\mathrm{T}} + UA_J^{-1}\sigma_\varepsilon^{-4}(U^{\mathrm{T}}U)A_J^{-1}U^{\mathrm{T}}\Big), \quad \text{with} \quad A_J \equiv I_2 + \sigma_\varepsilon^{-2}U^{\mathrm{T}}U.$$

Since $A_J^{-1} = O(J^{-1})$ and $U^{\mathrm{T}}U = O(J)$, the corrections are $O(J^{-1})$. Thus,

$$(\Sigma^{-2})_{j,j} = \sigma_\varepsilon^{-4}\{1 + O(J^{-1})\} \to \sigma_\varepsilon^{-4} \quad \text{uniformly in} \quad j.$$

Substituting these orders into

$$\sigma_{W,j}^2 = \sigma_f^2\big[(\Sigma^{-1}\beta)_j\big]^2 + \sigma_\varepsilon^2(\Sigma^{-2})_{j,j}$$

yields $\sigma_{W,j}^2 = \sigma_\varepsilon^{-2} + O(J^{-1})$ uniformly in j.

If $(f_z, \varepsilon_{j,z})$ are Gaussian, then $W_{z,j} \sim \mathcal{N}(0, \sigma_{W,j}^2)$ and

$$f_{W,j}(0) = \frac{1}{\sqrt{2\pi}\sigma_{W,j}} \to \frac{\sigma_\varepsilon}{\sqrt{2\pi}} \quad \text{uniformly in} \quad j.$$

The proof of the lemma is complete. ■

This result formalizes the intuition that as the cross-section expands, the factor-induced corrections to Σ^{-1} become negligible: the pseudo-inverse behaves almost like a scaled identity, and $f_W(0)$—the density at zero governing the linearization of the sign probability—is determined primarily by the idiosyncratic variance σ_ε^2.

Step 4. Letting $\Theta_1 \equiv \lim_{J\to\infty} \sum_{j=1}^{J} \mu_j$ as in (the first part of) expression (6), what is left to show is to establish (the second part of) expression (6) using population moments of (γ, β). Then, we can write expression (20) as

$$p(J) = \frac{1}{2} + \frac{C_1}{J} + O(J^{-2}) \quad \text{with} \quad C_1 = f_W(0) \cdot \Theta_1.$$

To find the closed-form expression for Θ_1, we define

$$r_J \equiv \frac{1}{J}\mathbf{1}^{\mathrm{T}}U = \begin{bmatrix} \mathbb{E}_J[\gamma] & \sigma_f \mathbb{E}_J[\beta] \end{bmatrix} \quad \text{and} \quad S_J \equiv \frac{1}{J}U^{\mathrm{T}}U = \begin{bmatrix} \mathbb{E}_J[\gamma^2] & \sigma_f \mathbb{E}_J[\gamma\beta] \\ \sigma_f \mathbb{E}_J[\gamma\beta] & \sigma_f^2 \mathbb{E}_J[\beta^2] \end{bmatrix}.$$

By the law of large numbers, $S_J \xrightarrow{\text{a.s.}} S$ and $r_J \xrightarrow{\text{a.s.}} r$, with

$$r \equiv \begin{bmatrix} \mathbb{E}[\gamma] & \sigma_f \mathbb{E}[\beta] \end{bmatrix} \quad \text{and} \quad S \equiv \begin{bmatrix} \mathbb{E}[\gamma^2] & \sigma_f \mathbb{E}[\gamma\beta] \\ \sigma_f \mathbb{E}[\gamma\beta] & \sigma_f^2 \mathbb{E}[\beta^2] \end{bmatrix}.$$

Then, Θ_1 admits the following expression.

Lemma 5 (Constant Θ_1) *The constant Θ satisfies:*

$$\Theta_1 = r^{\mathrm{T}} S^{-1} e_1 = \frac{\mathbb{E}[\gamma]\mathbb{E}[\beta^2] - \mathbb{E}[\beta]\mathbb{E}[\gamma\beta]}{\mathbb{E}[\gamma^2]\mathbb{E}[\beta^2] - \mathbb{E}[\gamma\beta]^2}.$$

Proof of Lemma 5. Observe that we have:

$$\Sigma^{-1}U = (\sigma_\varepsilon^2 I_J + UU^{\mathrm{T}})^{-1}U = U(\sigma_\varepsilon^2 I_2 + U^{\mathrm{T}}U)^{-1}.$$

Since $\gamma = Uv_1$,

$$\sum_{j=1}^{J} \mu_j = \mathbf{1}^{\mathrm{T}}\Sigma^{-1}\gamma = Jr_J^{\mathrm{T}}\left(\sigma_\varepsilon^2 I_2 + JS_J\right)^{-1} v_1 = r_J^{\mathrm{T}} S_J^{-1} v_1 + O(J^{-1}) \xrightarrow{\text{a.s.}} r^{\mathrm{T}} S^{-1} v_1.$$

Since $\det(S) = \sigma_f^2 \left(\mathbb{E}[\gamma^2]\mathbb{E}[\beta^2] - \mathbb{E}[\gamma\beta]^2\right) \neq 0$ holds (as $\sigma_f^2 > 0$ and β and γ are not colinear), we have:

$$S^{-1} = \frac{1}{\det(S)} \begin{bmatrix} \sigma_f^2 \mathbb{E}[\beta^2] & -\sigma_f \mathbb{E}[\gamma\beta] \\ -\sigma_f \mathbb{E}[\gamma\beta] & \mathbb{E}[\gamma^2] \end{bmatrix}.$$

Then, we obtain:

$$\Theta_1 = r^{\mathrm{T}} S^{-1} v_1 = \frac{\mathbb{E}[\gamma] \cdot \sigma_f^2 \mathbb{E}[\beta^2] + \sigma_f \mathbb{E}[\beta] \cdot (-\sigma_f \mathbb{E}[\gamma\beta])}{\det(S)} = \frac{\mathbb{E}[\gamma]\mathbb{E}[\beta^2] - \mathbb{E}[\beta]\mathbb{E}[\gamma\beta]}{\mathbb{E}[\gamma^2]\mathbb{E}[\beta^2] - \mathbb{E}[\gamma\beta]^2},$$

as desired. ■

This completes the proof of Theorem 2. Summarizing, the pseudo-inverse Y^+ acts asymptotically like a symmetric linear transformation applied to the noise and factor components, perturbed by a small deterministic mean shift of order $1/J$. The symmetry of the dominant stochastic term drives the limiting fraction of positive entries to one-half, while the deterministic correction produces the J^{-1} deviation summarized by the constant C_1.

A.3 Section 7

Proof of Proposition 3. The least-squares objective is

$$Q(S) = \|\Delta A_i - SG\|^2.$$

Differentiating with respect to S and setting the first-order condition to zero gives

$$-2(\Delta A_i - SG)G^{\mathrm{T}} = 0, \quad \text{that is,} \quad \Delta A_i G^{\mathrm{T}} = S(GG^{\mathrm{T}}).$$

If G has full row rank, then GG^{T} is invertible. Thus, the unique solution is

$$\hat{S}_i = \Delta A_i G^{\mathrm{T}} (GG^{\mathrm{T}})^{-1}.$$

Since $G^+ = G^{\mathrm{T}}(GG^{\mathrm{T}})^{-1}$ when G has full row rank, we have $\hat{S}_i = \Delta A_i G^+$.

Unbiasedness follows from assumed exogeneity of supply shocks,

$$\begin{aligned}\mathbb{E}[\hat{S}_i \mid G] &= \mathbb{E}[(S_i G + U_i)G^{\mathrm{T}}(GG^{\mathrm{T}})^{-1} \mid G] \\ &= S_i GG^{\mathrm{T}}(GG^{\mathrm{T}})^{-1} + \mathbb{E}[U_i G^{\mathrm{T}}(GG^{\mathrm{T}})^{-1} \mid G] \\ &= S_i.\end{aligned}$$

For consistency, assume $\frac{1}{N}GG^{\mathrm{T}} \to Q \succ 0$ for some positive definite Q. Further let $\mathbb{E}[\|U_i\|^2] < \infty$, and $\mathbb{E}[U_i \mid G] = 0$. Then

$$\hat{S}_i - S_i = U_i G^{\mathrm{T}}(GG^{\mathrm{T}})^{-1} = \left(\tfrac{1}{N}U_i G^{\mathrm{T}}\right)\left(\tfrac{1}{N}GG^{\mathrm{T}}\right)^{-1} \xrightarrow{p} 0,$$

by the law of large numbers for the cross-experiment averages. Thus $\hat{S}_i$ is consistent. ■

Proof of Proposition 4. Similarly to the proof of Proposition 4, the first-order condition is $S_i(GG^{\mathrm{T}}) = \Delta A_i G^{\mathrm{T}}$, from which we obtain $S_i G = \Delta A_i$. Multiplying G^+ from the right, we obtain $S_i P_G = \Delta A_i G^+$ as in the main text. Since this is a particular solution, the general solution can be written as

$$S_i = \Delta A_i G^+ + B_i(I_J - P_G),$$

where $B_i \in \mathbb{R}^{J \times J}$ is an arbitrary matrix and I_J is the identity matrix. ■

B Online Appendix

The Online Appendix is structured as follows. Appendix B.1 contains proofs for Section 3 and an example where redundant assets cause discontinuous demand. Appendix B.2 complements Section 4.3 by providing conditions under which Y^+ has the wrong sign for each state (Proposition 5), analogous to the asset-specific conditions in Proposition 2. Appendix B.3 extends Theorem 2 to a multifactor setting and offers further technical intuition on the constant Θ_1. Appendix B.4 supplements Section 6 by deriving asset demands and elasticities, illustrating state price variations, and discussing consumption implications. Finally, Appendix B.5 applies the incomplete identification result (Proposition 4) to green asset supply shocks in our illustrative example economy.

B.1 Section 3

Proof of Proposition 1. For the first statement, let $a^{*i} \in \mathcal{A}^i$ be a solution to (1). For ease of exposition, we allow 0 to be in the domain of u^i (this is not essential). Suppose to the contrary that there is an unbounded arbitrage opportunity. Since u^i is strictly increasing, there exists $m > 0$ such that

$$U^i(a^{*i}) < (1-\delta^i)u^i(e_0^i + p \cdot e^i) + \delta^i \pi_z u^i(m) + \delta^i(1-\pi_z)u^i(0) \text{ for some } z$$

and

$$U^i(a^{*i}) < (1-\delta^i)u^i(e_0^i + p \cdot e^i + m) + \delta^i u^i(0),$$

where $e^i \equiv (e_j^i)_{j=1}^J$. Since there is an unbounded arbitrage opportunity, for this $m > 0$, there exists $a^i \in \mathcal{A}^i$ such that either (i) $p \cdot a^i \leq 0$, $Y^{\mathrm{T}} a^i \geq 0$, and $(Y^{\mathrm{T}} a^i)_z \geq m$, in which case

$$U^i(a^{*i}) < (1-\delta^i)u^i(e_0^i + p \cdot e^i) + \delta^i \pi_z u^i(m) + \delta^i(1-\pi_z)u^i(0) \leq U^i(a^i)$$

or (ii) $p \cdot a^i \leq -m$ and $Y^{\mathrm{T}} a^i \geq 0$, in which case

$$U^i(a^{*i}) < (1-\delta^i)u^i(e^i_0 + p \cdot e^i + m) + \delta^i u^i(0) \leq U^i(a^i).$$

In either way, $a^{*i} \in \mathcal{A}^i$ does not solve (1), a contradiction.

For the second statement, since there is no unbounded arbitrage opportunity, there exists $m > 0$ such that, for any $a^i \in \mathcal{A}^i$,

$$U^i(a^i) < (1-\delta^i)u^i(e^i_0 + p \cdot e^i + m) + \delta^i u^i(m).$$

Thus, we obtain:

$$\sup_{a^i \in \mathcal{A}^i} U^i(a^i) \leq (1-\delta^i)u^i(e^i_0 + p \cdot e^i + m) + \delta^i u^i(m) < \infty.$$

Then, there exists a sequence $(a^{n,i})_{n\in\mathbb{N}}$ from $\mathcal{A}^i$ such that

$$\sup_{a^i \in \mathcal{A}^i} U^i(a^i) - \frac{1}{n} < U^i(a^{n,i}) \leq \sup_{a^i \in \mathcal{A}^i} U^i(a^i) < \infty \text{ for all } n \in \mathbb{N}.$$

Since $\sup_{a^i \in \mathcal{A}^i} U^i(a^i) < \infty$, it follows that

$$\sup_{n\in\mathbb{N}} |a^{n,i}_j| < \infty \text{ for all } j \in \{1, \ldots, J\}.$$

Since $\mathcal{A}^i$ is closed, it follows that there exists a convergent subsequence $(a^{n_k,i})_{k\in\mathbb{N}}$ of $(a^{n,i})_{n\in\mathbb{N}}$ such that $a^{n_k,i} \to a^{*i} \in \mathcal{A}^i$. Since U^i is continuous, it follows that

$$U^i(a^{*i}) = \sup_{a^i \in \mathcal{A}^i} U^i(a^i),$$

as desired. ■

Proof of Lemma 1. Suppose the conditions in the statement of the lemma. The proof consists of seven steps. First, for each $i \in I_0$, we define a subset M^i of $\mathbb{R}^{Z+1}$:

$$M^i \equiv \{(-p \cdot a^i, Y^{\mathrm{T}} a^i) \in \mathbb{R}^{Z+1} \mid a^i \in \mathcal{A}^i\}.$$

Then, for each $i \in I_0$, since investor i does not have an unbounded arbitrage opportunity, it follows that

$$M^i \cap \mathbb{R}_+^{Z+1} = \{0\}.$$

Note that $\mathbb{R}_+^{Z+1}$ is a closed convex cone in $\mathbb{R}^{Z+1}$ and does not contain any linear subspace other than $\{0\}$.

Second, let

$$M \equiv \bigcup_{i \in I_0} M^i.$$

It follows from the assumption

$$\mathbb{R}^J = \bigcup_{i \in I_0} \mathcal{A}^i$$

that

$$M = \{(-p \cdot a^i, Y^{\mathrm{T}} a^i) \in \mathbb{R}^{Z+1} \mid a^i \in \mathbb{R}^J\}$$

is a linear subspace.

Third, since

$$M \cap \mathbb{R}_+^{Z+1} = \{0\},$$

it follows from the separating hyperplane theorem (which is referred to as "Linear separation of Cones" in Duffie (2001)), there exists $\overline{q} \in \mathbb{R}^{Z+1} \setminus \{0\}$ such that

$$\overline{q} \cdot t < \overline{q} \cdot x \text{ for all } t \in M \text{ and } x \in \mathbb{R}_+^{Z+1}.$$

Fourth, we show that $\overline{q} \in \mathbb{R}_{++}^{Z+1}$. Since $0 \in M$, it follows that

$$0 = \overline{q} \cdot 0 < \overline{q} \cdot x \text{ for all } x \in \mathbb{R}_+^{Z+1}.$$

Taking x as standard unit vectors in $\mathbb{R}_{++}^{Z+1}$ yields $\overline{q}_z > 0$ for all z.

Fifth, we show that

$$0 = \overline{q} \cdot t \text{ for all } t \in M.$$

Suppose to the contrary that $0 \neq \overline{q} \cdot t$ for some $t \in M$. Since M is a linear subspace, we can assume, without loss, that

$$\overline{q} \cdot t > 0.$$

However, this leads to a contradiction because, for any given $x \in \mathbb{R}^{Z+1}_{++}$, there exists $\lambda \in \mathbb{R}$ such that $\lambda t \in M$ and

$$\overline{q} \cdot x \leq \lambda(\overline{q} \cdot t) = \overline{q} \cdot (\lambda t).$$

Sixth, we show that

$$\overline{q}^{\mathrm{T}} \begin{bmatrix} -p^{\mathrm{T}} \\ Y^{\mathrm{T}} \end{bmatrix} = 0.$$

It follows from the fifth step that

$$\overline{q}^{\mathrm{T}} \begin{bmatrix} -p^{\mathrm{T}} \\ Y^{\mathrm{T}} \end{bmatrix} a = 0 \text{ for all } a \in \mathbb{R}^J = \bigcup_{i \in I_0} \mathcal{A}^i.$$

If

$$\overline{q}^{\mathrm{T}} \begin{bmatrix} -p^{\mathrm{T}} \\ Y^{\mathrm{T}} \end{bmatrix} \neq 0,$$

then letting

$$a = \left(\overline{q}^{\mathrm{T}} \begin{bmatrix} -p^{\mathrm{T}} \\ Y^{\mathrm{T}} \end{bmatrix} \right)^{\mathrm{T}} \in \mathbb{R}^J = \bigcup_i \mathcal{A}^i$$

yields

$$\overline{q}^{T} \begin{bmatrix} -p^{\mathrm{T}} \\ Y^{\mathrm{T}} \end{bmatrix} a > 0,$$

a contradiction.

Seventh, then, denoting by

$$\overline{q} = (q_0, q_{-0}),$$

we have

$$\bar{q}_0 p^{\mathrm{T}} = q_{-0}^{\mathrm{T}} Y^{\mathrm{T}}, \text{ that is, } p = Y\frac{q_{-0}}{q_0}.$$

Letting $q = \frac{q_{-0}}{q_0} \in \mathbb{R}^Z_{++}$, we finally obtain

$$p = Yq,$$

as desired. ■

B.1.1 Arbitrage and Discountinuous Demands

This example illustrates an example in which an asset demand function exhibits discontinuity in the presence of redundant assets.

Example 2 (Discontinuous demand functions) *Suppose there are two states of the world at date 1, and three assets. Given some $\epsilon \in (0,1)$, let a cash flow matrix Y be given by*

$$\begin{bmatrix} \frac{1}{2}(1+\epsilon) & \frac{1}{2}(1-\epsilon) \\ \frac{1}{2}(1-\epsilon) & \frac{1}{2}(1+\epsilon) \\ 1 & 1 \end{bmatrix}.$$

Now consider the demand functions for some investor i with continuous utility function U^i.

(i) Suppose $\mathcal{A}^i = \mathbb{R}^3$. The absence of unbounded arbitrage requires that $p_3 = p_1 + p_2$. Given this restriction on prices, well-defined demand functions exist for all three assets, with the investor taking weakly positive quantities in all three assets. Now suppose that, starting from an initial benchmark where no arbitrage pricing holds, p_3 increases slightly. Then, investor i's problem (1) is no longer well-defined, and well-defined demand functions no longer exist.

(ii) Suppose instead that investor i faces the short-sale constraint $a^i_j \geq -\chi$ for some $\chi > 0$. Given $p_3 = p_1 + p_2$, well-defined demand functions still exist for all three assets, with the investor taking weakly positive quantities in all three assets. Now suppose that p_3 increases slightly. Then it is optimal for the investor to jump to

a portfolio allocation where $a_3^i = -\chi$. This can trigger discontinuities in optimal demand.

B.2 Section 4.3

We remark that we can also provide conditions under which Y^+ has a wrong sign for each state (i.e., row).

Proposition 5 *Under the following two properties, each row of Y^+ contains at least one negative element: for each $z \in \{1, \ldots, Z\}$, there exists at least one $j \in \{1, \ldots, J\}$ such that $(Y^+)_{z,j} < 0$.*

(i) Each row of Y has at least two strictly positive elements.

(ii) Conical Independence: no column vector $y(z)$ of Y can be written as a non-negative linear combination of the other column vectors of Y: for any $z \in \{1, \ldots, Z\}$, there exists no $(\alpha_{z'})_{z' \neq z} \in \mathbb{R}_+^{Z-1}$ such that

$$y(z) = \sum_{z' \neq z} \alpha_{z'} y(z').$$

Before proving Proposition 5, we discuss its assumptions. Property (i) states that assets typically pay off in multiple states, ruling out only the knife-edge case of Arrow securities. Property (ii) is a weak linear independence requirement: it rules out perfectly redundant states whose payoffs can be exactly replicated by combinations of other states. In the special case in which $J = Z$, property (ii) is automatically satisfied because the assumption that $\text{rank}(Y) = J$ implies that the columns of Y are linearly independent. These properties hold in virtually all realistic asset markets.

Proof of Proposition 5. Let $y(z)$ be the z-th column of Y. Let y_k^+ be the k-th row of Y^+. Suppose to the contradiction that there exists a row k such that $y_k^+ \geq 0$ element-by-element.

Consider the projection matrix $P = Y^+Y$. The entries are given by $P_{kz} = y_k^+ \cdot y(z)$. It follows from $y_k^+ \geq 0$ and $y(z) \geq 0$ that

$$P_{kz} \geq 0 \quad \text{for all} \quad z \in \{1, \ldots, Z\}.$$

The columns of Y span the range of Y. The projection matrix P acts as the identity on the row space of Y^T, which implies $YP = Y$. Writing this column-wise for vector $y(z)$, for each $z \in \{1, \ldots, Z\}$, it follows from $y(z) = YP_{\cdot,z}$ that

$$y(z) = \sum_{k=1}^{Z} P_{kz} y(k), \quad \text{that is,} \quad (1 - P_{zz})y(z) = \sum_{k \neq z} P_{kz} y(k).$$

Since P is a projection matrix, $P_{zz} \leq 1$.

If $P_{zz} < 1$, then we have

$$y(z) = \sum_{k \neq z} \frac{P_{kz}}{1 - P_{zz}} y(k),$$

which is a contradiction to property (ii).

Thus, suppose that $P_{zz} = 1$. Then, $\sum_k P_{zk}^2 = P_{zz}$ implies $P_{zk} = 0$ for all $k \neq z$. This implies

$$P_{zk} = y_z^+ \cdot y(k) = 0 \quad \text{for all} \quad k \neq z.$$

Since $y_z^+ \geq 0$ and $y(k) \geq 0$, let

$$S = \{m \in \{1, \ldots, J\} \mid (y_z^+)_m > 0\}, \quad \text{where} \quad (y_z^+)_m = (Y^+)_{z,m}.$$

The set S is not empty because $y_z^+ \cdot y(z) = P_{zz} = 1$. For all $k \neq z$, and for all $m \in S$, we must have $0 = y_m(k) (= Y_{m,k})$. Take any index $m \in S$. The row m of matrix Y has a value of 0 in every column $k \neq z$. Therefore, row m contains at most one strictly positive element (potentially at column z). This contradicts property (i). ■

B.3 Section 5 (Theorem 2)

B.3.1 Multifactor Extension of Theorem 2

The structure of the proof for a finite number of factors K is identical to the one-factor argument. Consider the K-factor model

$$y_{j,z} = \alpha_j + \sum_{k=1}^{K} \beta_j^{(k)} f_z^{(k)} + \varepsilon_{j,z} = \underbrace{\left(\alpha_j + \sum_{k=1}^{K} \beta_j^{(k)} \bar{f}^{(k)}\right)}_{\gamma_j} + \sum_{k=1}^{K} \beta_j^{(k)} \left(f_z^{(k)} - \bar{f}^{(k)}\right) + \varepsilon_{j,z},$$

where $\bar{f}^{(k)} = \mathbb{E}[f_z^{(k)}]$. Assumptions (A1)-(A3) are naturally extended to K factors: that is, the vectors $(\beta_j^{(k)})_{k=1}^K$ replace β_j and the factors $(f_z^{(k)})_{k=1}^K$ replace f_z; all factors are mutually independent and independent of errors.

Each factor adds an additional "spike" to the covariance matrix,

$$\Sigma = \gamma\gamma^{\mathrm{T}} + \sum_{k=1}^{K} \sigma_{f,k}^2 \beta^{(k)} \beta^{(k)\mathrm{T}} + \sigma_\varepsilon^2 I_J = \sigma_\varepsilon^2 I_J + UU^{\mathrm{T}} \quad \text{with} \quad U \equiv \begin{bmatrix} \gamma & \sigma_{f,1}\beta^{(1)} & \cdots & \sigma_{f,K}\beta^{(K)} \end{bmatrix},$$

but the key asymptotic properties remain unchanged. The deterministic mean shift $\mu_j \equiv (\Sigma^{-1}\gamma)_j$ is still of order J^{-1}, and the random fluctuation

$$W_{z,j} \equiv \sum_{k=1}^{K} (f_z^{(k)} - \bar{f}^{(k)})(\Sigma^{-1}\beta^{(k)})_j + (\Sigma^{-1}\varepsilon_z)_j$$

remains symmetric around zero. Consequently, the linear expansion of the sign probability and the $O(J^{-1})$ convergence rate carry over verbatim. The only new element is the form of the constant Θ_K, which now depends on the $(K+1) \times (K+1)$ population moment matrix of vectors U.

In general, letting $S = \mathbb{E}[UU^{\mathrm{T}}]$ and

$$r = \mathbb{E}[U] = \begin{bmatrix} \mathbb{E}[\gamma] & \sigma_{f,1}\mathbb{E}[\beta^{(1)}] & \cdots & \sigma_{f,K}\mathbb{E}[\beta^{(K)}] \end{bmatrix},$$

we have

$$\Theta_K = r^{\mathrm{T}} S^{-1} v_1.$$

When $K = 1$, this expression reduces to the right-most side of expression (6). We summarize the K-factor extension as follows.

Corollary 1 (K-Factor Extension) *In the K-factor model, the same asymptotic result holds:*

$$p(J) = \frac{1}{2} + \frac{C_K}{J} + O(J^{-2}) \quad a.s., \qquad \text{with} \quad C_K = f_W(0) \cdot \Theta_K \quad \text{and} \quad \Theta_K = r^{\mathrm{T}} S^{-1} v_1.$$

Consequently,

$$\operatorname*{plim}_{J \to \infty} p(J) = \frac{1}{2}.$$

B.3.2 Remark on the Constant Θ_1

When $\sigma_f^2 \|\beta\|^2 \ll \|\gamma\|^2$, or equivalently, when $\sigma_f^2 \mathbb{E}[\beta^2] \ll \mathbb{E}[\gamma^2]$ in the limit as $J \to \infty$, we can circumvent arguments in Step 3.[10] Indeed, Σ admits the following simpler approximation:

$$\Sigma \approx \gamma\gamma^{\mathrm{T}} + \sigma_\epsilon^2 I_J.$$

Applying the Woodbury identity, we obtain:

$$\Sigma^{-1} \approx \frac{1}{\sigma_\epsilon^2} \left(I_J - \frac{\gamma\gamma^{\mathrm{T}}}{\sigma_\epsilon^2 + \|\gamma\|^2} \right). \tag{22}$$

Then, operating γ on expression (22) from the right, we obtain:

$$\Sigma^{-1}\gamma \approx \frac{\gamma}{\sigma_\epsilon^2 + \|\gamma\|^2}.$$

When J is large, since

$$\mu_j = (\Sigma^{-1}\gamma)_j \approx \frac{\gamma_j}{\sigma_\epsilon^2 + \|\gamma\|^2} \approx \frac{\gamma_j}{J \cdot \mathbb{E}[\gamma^2]},$$

it follows that

$$\Theta_1 \approx \frac{\mathbb{E}[\gamma]}{\mathbb{E}[\gamma^2]}.$$

[10] For the K-factor case, $\sigma_{f,k}^2 \|\beta^{(k)}\|^2 \ll \|\gamma\|^2$ for all k.

Similarly, operating a scalar σ_ϵ^2 on expression (22) from the right, the variance of the stochastic term is dominated by

$$(\Sigma^{-1})_{j,j}\sigma_\epsilon^2 \approx \sigma_\epsilon^{-2} \cdot \sigma_\epsilon^2 = 1,$$

yielding

$$\sigma_W \approx \frac{1}{\sigma_\epsilon}.$$

B.4 Section 6

Equation (10). For ease of presentation, we derive the demand functions a_g and a_r directly from the representative agent's portfolio choice problem:[11]

$$\max_{a_g,a_r} (1-\delta)u(E_0 - p_g(a_g - E_g) - p_r(a_r - E_r)) + \delta\pi_g u(y_g(g)a_g + y_r(g)a_r) + \delta\pi_r u(y_g(r)a_g + y_r(r)a_r).$$

After substituting the payoff matrix Y into the utility function, the first-order conditions are:

$$(1-\delta)\frac{p_g}{E_0 - p_g(a_g - E_g) - p_r(a_r - E_r)} = \delta\pi_g \frac{1+\epsilon}{(1+\epsilon)a_g + (1-\epsilon)a_r} + \delta\pi_r \frac{1-\epsilon}{(1-\epsilon)a_g + (1+\epsilon)a_r}; \tag{23}$$

$$(1-\delta)\frac{p_r}{E_0 - p_g(a_g - E_g) - p_r(a_r - E_r)} = \delta\pi_g \frac{1-\epsilon}{(1+\epsilon)a_g + (1-\epsilon)a_r} + \delta\pi_r \frac{1+\epsilon}{(1-\epsilon)a_g + (1+\epsilon)a_r}. \tag{24}$$

Then, since $\pi_g = 1 - \pi_r$, the representative agent's demand functions are:

$$a_g(p_g, p_r) = \delta\frac{(E_0 + p_g E_g + p_r E_r)\left((1-\epsilon^2)p_g - ((1+\epsilon)^2 - 4\epsilon\pi_r)p_r\right)}{(p_g - p_r)^2 - (p_g + p_r)^2\epsilon^2};$$

$$a_r(p_g, p_r) = \delta\frac{(E_0 + p_g E_g + p_r E_r)\left((1-\epsilon^2)p_r - ((1-\epsilon)^2 + 4\epsilon\pi_r)p_g\right)}{(p_g - p_r)^2 - (p_g + p_r)^2\epsilon^2}.$$

[11] Alternatively, one can derive the demand functions a_g and a_r from the consumption functions c_g and c_r.

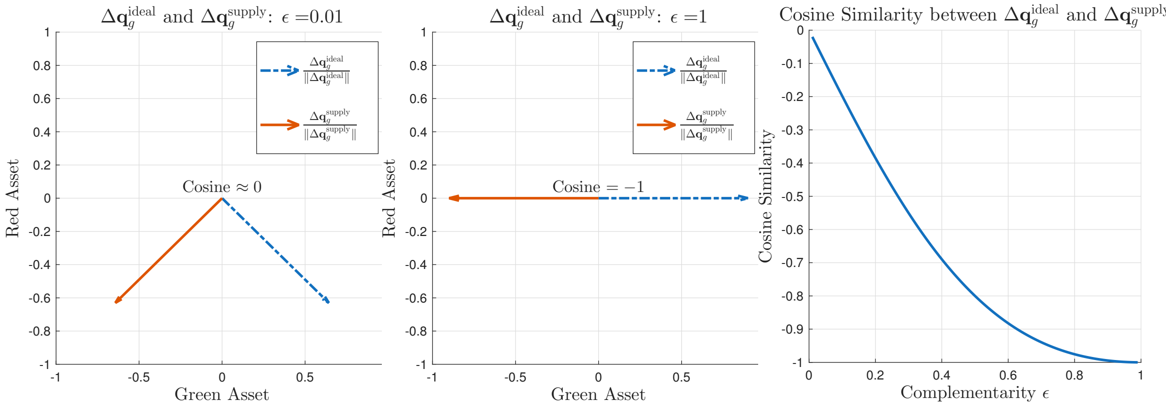

Figure 3: State Price Variations $\Delta \mathbf{q}_g^{\text{ideal}}$ and $\Delta \mathbf{q}_g^{\text{supply}}$. Parameters: $\pi_g = \pi_r = \frac{1}{2}$.

Elasticities. On the one hand,

$$\mathcal{E}^{\text{ideal}} \equiv -\frac{\partial a_g(p_g, p_r)}{\partial p_g} \frac{p_g}{a_g}$$
$$= (1 + (1 - 2\pi_r)\epsilon) \frac{(1-\epsilon)^2 + 4\epsilon\pi_r(1-\delta\epsilon) + 4\delta\epsilon^2\pi_r^2}{8\pi_r(1-\pi_r)\epsilon^2}.$$

On the other hand, since

$$-\frac{\frac{da_g}{ds_g}}{\frac{dp_g}{ds_g}} = -\frac{\partial a_g}{\partial p_g} - \frac{\partial a_g}{\partial p_r} \frac{\frac{dp_r}{ds_g}}{\frac{dp_g}{ds_g}}$$
$$= \frac{1-\delta}{\delta} \frac{(1-\epsilon)^2 + 4\epsilon\pi_r(1-\delta\epsilon) + 4\delta\epsilon^2\pi_r^2}{4\epsilon^2\pi_r(1-\pi_r)} - \frac{1-\delta}{\delta} \frac{(1-\epsilon^2) + 4\delta\epsilon^2\pi_r(1-\pi_r)}{4\epsilon^2\pi_r(1-\pi_r)} \frac{1-\epsilon^2}{(1+\epsilon)^2 - 4\epsilon\pi_r}$$
$$= 2\frac{1-\delta}{\delta} \frac{2 - \delta(1 + (1-2\pi_r)\epsilon)}{(1+\epsilon)^2 - 4\epsilon\pi_r},$$

it follows that

$$\mathcal{E}^{\text{supply}} \equiv -\frac{\frac{da_g}{ds_g}}{\frac{dp_g}{ds_g}} \frac{p_g}{a_g}$$
$$= (1 + (1-2\pi_r)\epsilon) \frac{2 - \delta(1 + (1-2\pi_r)\epsilon)}{(1+\epsilon)^2 - 4\epsilon\pi_r}.$$

B.4.1 State Price Variations in the General Equilibrium Example

Figure 3 compare the state price variations $\Delta \mathbf{q}_g^{\text{ideal}}$ and $\Delta \mathbf{q}_g^{\text{supply}}$ in our illustrative example. Note that while the state price variation $\Delta \mathbf{q}_g^{\text{ideal}}$ is given by (8) , the state price variation $\Delta \mathbf{q}_g^{\text{supply}}$ is given by:

$$\Delta \mathbf{q}_g^{\text{supply}} = \frac{\partial}{\partial s_g} \begin{bmatrix} q_g \\ q_r \end{bmatrix} \Bigg|_{s_g \to 0} = -\frac{\delta}{1-\delta} \begin{bmatrix} \pi_g \frac{1+\epsilon}{2} \\ \pi_r \frac{1-\epsilon}{2} \end{bmatrix} (<0).$$

The left and central panels depict the state price variations for a fixed ϵ. The right panel depicts the cosine similarity between $\Delta \mathbf{q}_g^{\text{ideal}}$ and $\Delta \mathbf{q}_g^{\text{supply}}$.

B.4.2 Implications for Demand

The fact that the supply shock generates the wrong type of state price variation dramatically affects the observed demand response. We illustrate this effect by computing the response of the consumption ratio c_g/c_r to both the ideal experiment and the supply shock. Given log utility, it follows from the first-order conditions (7) that the relative consumption process satisfies:

$$\frac{c_g}{c_r} = \frac{\pi_g}{\pi_r} \frac{q_r}{q_g}. \tag{25}$$

Relative consumption in turn determines the desired holdings of green and red assets.

Consider first the ideal experiment with a pure price shock. Differentiating the relative consumption with respect to p_g and evaluating in the limit $s_g \to 0$ yields:

$$-\frac{\partial}{\partial p_g} \left(\frac{c_g}{c_r} \right) \Bigg|_{s_g \to 0} = \frac{1-\delta}{\delta} \frac{(1-\epsilon)\pi_g + (1+\epsilon)\pi_r}{2\pi_g \pi_r \epsilon}. \tag{26}$$

This derivative diverges to infinity as $\epsilon \to 0$. As the two assets are perfect substitutes in this limit, a small *price* shock triggers a rapid reallocation from green to red assets.

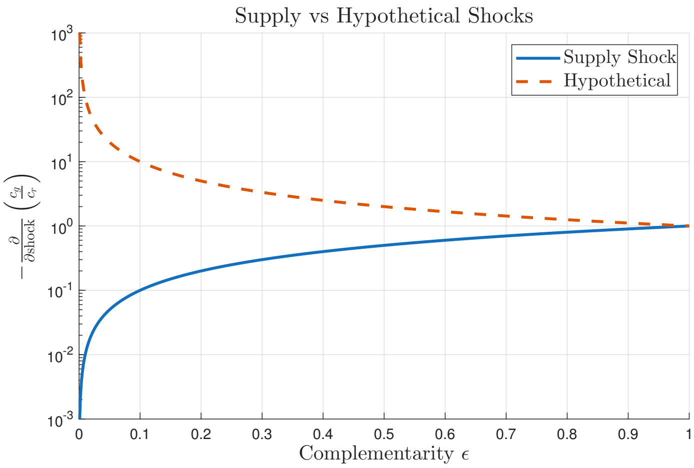

Figure 4: Optimal change in consumption ratio c_g/c_r on log scale. Parameters: $\pi_g = \pi_r = \frac{1}{2}$ and $\delta = \frac{2}{3}$.

Next, consider the response to the supply shock. In the limit as $s_g \to 0$,

$$\frac{\partial}{\partial s_g}\left(\frac{c_g}{c_r}\right)\bigg|_{s_g \to 0} = \epsilon. \tag{27}$$

which converges to *zero* in the limiting case of perfect substitutes as $\epsilon \to 0$. When the two assets are perfect substitutes, a supply shock has identical effects in both states. As such, it results in *zero* difference in the optimal consumption ratio across the two states.

Figure 4 depicts the optimal investor-level response to the hypothetical price shock (26) and the response to the supply shock (27) on log scale (Appendix B.4.3 provides the derivations of these expressions). The difference in responses diverges to infinity as $\epsilon \to 0$. The only point of overlap occurs when the two assets are both Arrow securities. In line with our theory, this is the case where there can be no spillovers across assets.

B.4.3 Derivations for Appendix B.4.2

Equation (26). Since the Arrow prices q can be expressed as a function of the asset prices p through $p = Yq$, the consumption ratio (25) can be written as:

$$\frac{c_g}{c_r} = \frac{\pi_g}{\pi_r} \frac{(1+\epsilon)p_r - (1-\epsilon)p_g}{(1+\epsilon)p_g - (1-\epsilon)p_r}.$$

Thus, differentiating it with respect to the price p_g, we have:

$$\frac{\partial}{\partial p_g}\left(\frac{c_g}{c_r}\right) = -\frac{\pi_g}{\pi_r} \frac{4\epsilon p_r}{((1+\epsilon)p_g - (1-\epsilon)p_r)^2}. \tag{28}$$

In contrast, substituting the Arrow prices (9) into $p = Yq$, we obtain:

$$p_g = \frac{1+\epsilon}{2}\pi_g \frac{\delta}{1-\delta} \frac{1}{1+\frac{1+\epsilon}{2}s_g} + \frac{1-\epsilon}{2}\pi_r \frac{\delta}{1-\delta} \frac{1}{1+\frac{1-\epsilon}{2}s_g}; \tag{29}$$

$$p_r = \frac{1-\epsilon}{2}\pi_g \frac{\delta}{1-\delta} \frac{1}{1+\frac{1+\epsilon}{2}s_g} + \frac{1+\epsilon}{2}\pi_r \frac{\delta}{1-\delta} \frac{1}{1+\frac{1-\epsilon}{2}s_g}. \tag{30}$$

Substituting the asset prices p at $s_g = 0$ into equation (28), we obtain equation (26). When $\delta = \frac{2}{3}$ and $\pi_g = \pi_r = \frac{1}{2}$, equation (26) reduces to:

$$-\left.\frac{\partial}{\partial p_g}\left(\frac{c_g}{c_r}\right)\right|_p = \frac{1}{\epsilon}.$$

Equation (27). Substituting the Arrow prices (9) into the consumption ratio (25) yields

$$\frac{c_g}{c_r} = \frac{1+\frac{1+\epsilon}{2}s_g}{1+\frac{1-\epsilon}{2}s_g}.$$

Thus, differentiating it with respect to the supply shock s_g, we obtain

$$\frac{\partial}{\partial s_g}\left(\frac{c_g}{c_r}\right) = \frac{\epsilon}{\left(1+\frac{1-\epsilon}{2}s_g\right)^2}.$$

In the limit as $s_g \to 0$, we get equation (27).

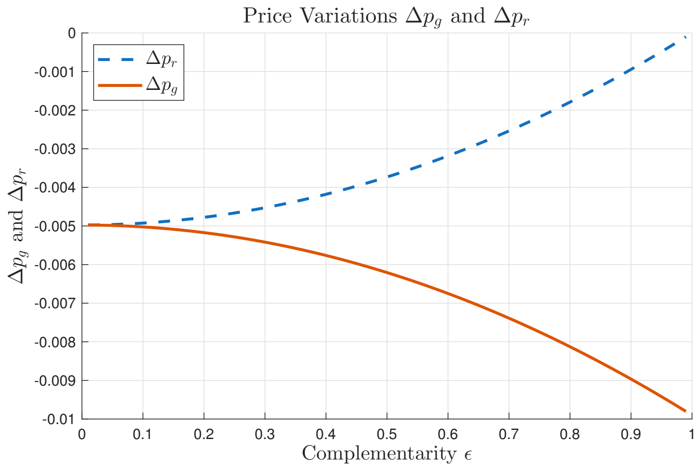

Figure 5: Price Changes Δp_g and Δp_r. Parameters: $\pi_g = \pi_r = \frac{1}{2}$, $\delta = \frac{2}{3}$, and $s_g = 0.01$.

B.5 Section 7.1

We now illustrate the incomplete identification result in the context of our example economy where the supply shock to the green asset is a single experiment. This creates a vector of observable price changes G and that of quantity changes ΔA for the representative investor:

$$G = \begin{bmatrix} \Delta p_g \\ \Delta p_r \end{bmatrix} \quad \text{and} \quad \Delta A = \begin{bmatrix} \Delta a_g \\ \Delta a_r \end{bmatrix} = \begin{bmatrix} s_g \\ 0 \end{bmatrix}.$$

Figure 5 illustrates the price changes G for varying complementarity ϵ. In the limit as $\epsilon \to 1$, the vector G reduces to the pure price change of the green asset.

We compare the theoretical asset-level substitution matrix

$$S = \begin{bmatrix} \frac{\partial a_g}{\partial p_g} & \frac{\partial a_g}{\partial p_r} \\ \frac{\partial a_r}{\partial p_g} & \frac{\partial a_r}{\partial p_r} \end{bmatrix} \tag{31}$$

and its least-square identification

$$\Delta A G^+ = \frac{s_g}{\|\Delta p\|^2} \begin{bmatrix} \Delta p_g & \Delta p_r \\ 0 & 0 \end{bmatrix}. \tag{32}$$

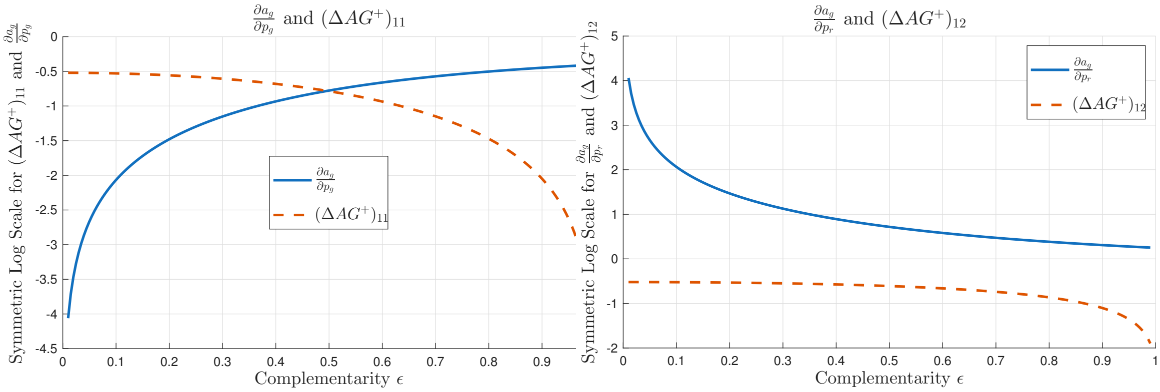

Figure 6: Comparison of S and ΔAG^{+}. Parameters: $\pi_g = \pi_r = \frac{1}{2}$ and $\delta = \frac{2}{3}$.

The left panel of Figure 6 compares the $(1, 1)$ element of S and ΔAG^{+} on the symmetric log scale, while the right panel of Figure 6 does the $(1, 2)$ element of S and ΔAG^{+}.[12] The right panel suggests that the signs are opposite.

B.5.1 Derivations

Equation (31). The first-order conditions (23) and (24) yield the equilibrium prices p_g and p_r. When $(E_0, E_g, E_r) = (1, 1 + s_g, 1)$, p_g and p_r coincide with equations (29) and (30). We denote by $\overline{p}$ the initial equilibrium price vector (precisely, equations (29) and (30) with $s_g = 0$):

$$\overline{p}_g = \frac{1}{2}\frac{\delta}{1-\delta}\left(1 + (1 - 2\pi_r)\epsilon\right) \quad \text{and} \quad \overline{p}_r = \frac{1}{2}\frac{\delta}{1-\delta}\left(1 - (1 - 2\pi_r)\epsilon\right).$$

Then, the matrix

$$S = \begin{bmatrix} \frac{\partial a_g(\overline{p})}{\partial p_g} & \frac{\partial a_g(\overline{p})}{\partial p_r} \\ \frac{\partial a_r(\overline{p})}{\partial p_g} & \frac{\partial a_r(\overline{p})}{\partial p_r} \end{bmatrix}$$

[12] The symmetric log scale transforms any x into $\text{sign}(x)\log_{10}(1 + \log_e(10)|x|)$. Thus, it respects the sign of x, it is close to x when x is small, and it is an approximate logarithmic scale when x is large.

is given by:

$$
\begin{aligned}
\frac{\partial a_g(\overline{p})}{\partial p_g} &= -\frac{1-\delta}{\delta}\frac{(1-\epsilon)^2+4\epsilon\pi_r\,(1-\delta\epsilon)+4\delta\epsilon^2\pi_r^2}{4\epsilon^2\pi_r(1-\pi_r)};\\
\frac{\partial a_g(\overline{p})}{\partial p_r} &= \frac{1-\delta}{\delta}\frac{(1-\epsilon^2)+4\delta\epsilon^2\pi_r(1-\pi_r)}{4\epsilon^2\pi_r(1-\pi_r)};\\
\frac{\partial a_r(\overline{p})}{\partial p_g} &= \frac{1-\delta}{\delta}\frac{(1-\epsilon^2)+4\delta\epsilon^2\pi_r(1-\pi_r)}{4\epsilon^2\pi_r(1-\pi_r)}\left(=\frac{\partial a_g(\overline{p})}{p_r}\right);\\
\frac{\partial a_r(\overline{p})}{\partial p_r} &= -\frac{1-\delta}{\delta}\frac{(1+\epsilon)^2-4\epsilon\pi_r(1+\delta\epsilon)+4\delta\epsilon^2\pi_r^2}{4\epsilon^2\pi_r(1-\pi_r)}.
\end{aligned}
$$

Equation (32). Since the vector of price changes is given by

$$
G \equiv \begin{bmatrix}\Delta p_g\\ \Delta p_r\end{bmatrix} = \begin{bmatrix}p_g-\overline{p}_g\\ p_r-\overline{p}_r\end{bmatrix},
$$

its Moore-Penrose inverse is a 1×2 matrix $G^+ = (G^{\mathrm{T}}G)^{-1}G^{\mathrm{T}}$, that is,

$$
G^+ = \begin{bmatrix}\frac{p_g-\overline{p}_g}{(p_g-\overline{p}_g)^2+(p_r-\overline{p}_r)^2} & \frac{p_r-\overline{p}_r}{(p_g-\overline{p}_g)^2+(p_r-\overline{p}_r)^2}\end{bmatrix} = \frac{(\Delta p)^{\mathrm{T}}}{\|\Delta p\|^2},
$$

where

$$
\|\Delta p\|^2 = s_g^2\delta^2\frac{(1+\epsilon)^2(2+(1-\epsilon)s_g)^2(1+\epsilon^2)-4\epsilon\pi_r(1+\epsilon)^2(2+s_g(1-\epsilon))(2+s_g(1-\epsilon)\epsilon)+4\epsilon^2\pi_r^2\left(s_g\left(4+s_g-2(2+s_g)\epsilon^2+s_g\epsilon^4\right)\right)}{2(1-\delta)^2\left(4+s_g(4+(1-\epsilon^2)s_g)\right)^2}.
$$

Then, the least-square solution ΔAG^+ is:

$$
\Delta AG^+ = \frac{1}{\|\Delta p\|^2}\begin{bmatrix}s_g\\0\end{bmatrix}\begin{bmatrix}\Delta p_g & \Delta p_r\end{bmatrix} = \frac{s_g}{\|\Delta p\|^2}\begin{bmatrix}\Delta p_g & \Delta p_r\\ 0 & 0\end{bmatrix}.
$$