A Trilemma for Asset Demand Estimation*

Most recent version

William Fuchs† Satoshi Fukuda‡ Daniel Neuhann§

May 8, 2026

Abstract

We establish fundamental limits on demand analysis in asset markets. Given observational data, one cannot jointly maintain (i) no-arbitrage asset pricing, (ii) investor preferences over payoffs, and (iii) model-free identification of structural asset demand. These results follow from a decomposition of asset demand into fundamental preferences and a latent mapping from preferences to asset holdings. This mapping is fundamentally unobservable, varies with routine changes to expected asset payoffs, and links all asset-level demand functions into a non-separable system. As such, even perfectly exogenous supply shifters cannot produce the price variation required to estimate stable asset demand curves. We conclude that asset demand elasticities are contingent, model-specific constructs which reflect—rather than validate—a priori assumptions on investor behavior.

*We thank Dan Ackerberg, Yu An, Jonathan Berk, Olivier Darmouni, Aureo de Paula, Julie Fu, Zhiguo He, Andrey Ordin, and seminar participants at UCL, UT Austin, University of Zurich, 35th Annual Utah Winter Finance Conference, 7th Columbia Workshop in New Empirical Finance, 19th Jackson Hole Finance Group Conference, 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.

A Trilemma for Asset Demand Estimation*

Most recent version

William Fuchs† Satoshi Fukuda‡ Daniel Neuhann§

May 8, 2026

Abstract

We establish fundamental limits on demand analysis in asset markets. Given observational data, one cannot jointly maintain (i) no-arbitrage asset pricing, (ii) investor preferences over payoffs, and (iii) model-free identification of structural asset demand. These results follow from a decomposition of asset demand into fundamental preferences and a latent mapping from preferences to asset holdings. This mapping is fundamentally unobservable, varies with routine changes to expected asset payoffs, and links all asset-level demand functions into a non-separable system. As such, even perfectly exogenous supply shifters cannot produce the price variation required to estimate stable asset demand curves. We conclude that asset demand elasticities are contingent, model-specific constructs which reflect—rather than validate—a priori assumptions on investor behavior.

---

*We thank Dan Ackerberg, Yu An, Jonathan Berk, Olivier Darmouni, Aureo de Paula, Julie Fu, Zhiguo He, Andrey Ordin, and seminar participants at UCL, UT Austin, University of Zurich, 35th Annual Utah Winter Finance Conference, 7th Columbia Workshop in New Empirical Finance, 19th Jackson Hole Finance Group Conference, 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

How much do investors want to hold of a given asset, and how sensitive are their portfolio choices to the price? These questions lie at the heart of asset pricing. They determine how much prices move when central banks purchase bonds, when passive funds rebalance indices, or when financial intermediaries suffer shocks that force asset sales. An influential empirical literature has set out to answer them by estimating asset demand functions from data on portfolio holdings and prices.

A critical question for this literature is whether observed demand responses can be given a structural interpretation without relying on a fully specified equilibrium model of portfolio choice. If so, demand elasticities are invariant to the model used to estimate them and can be used to discriminate between different theories of investor behavior. If not, they are contingent, model-specific objects that reflect — rather than validate — a priori assumptions on investor behavior.

We provide a general theoretical analysis of this question. Our answer is that asset demand functions are not model-free empirical objects, and that structural modeling is unavoidable. Our results require only two foundational principles of asset pricing: that investors value assets for their payoffs, and that asset prices admit no arbitrage. As we discuss, these principles are difficult to discard without invalidating the basic premise of asset demand analysis.

Our results follow from a general decomposition of asset demand functions derived under preferences over payoffs and no arbitrage. Denote by Y the *payoff matrix* summarizing investor beliefs over the state-contingent payoffs of assets in the choice set. Then the matrix of asset demand slopes (in which each element is the derivative of asset demand with respect to a specific asset price) satisfies

$$\mathcal{A}^i = (Y^+)^{\mathrm{T}} \mathcal{D}^i Y^+.$$

In this decomposition, $\mathcal{D}^i$ is the investor's *fundamental demand function* for state-contingent payoffs and Y^+ is the Moore-Penrose pseudo-inverse of Y. This has a natural interpretation: since preferences are defined over payoffs, not assets directly, Y^+ maps demand for payoffs into the associated asset quantities.

This simple decomposition reveals two main challenges. First, the demand function for any individual asset is *commingled with those of all other assets*: because investors care about state-contingent consumption, the optimal quantity of any asset generically depends on the payoffs of all other assets through Y^+. This means that one cannot analyze demand for any given asset in isolation. Second, Y^+ is *latent and unidentifiable* from past data. Since the payoff matrix reflects investor beliefs about future payoffs, including resale prices, no finite sample of realized returns can pin it down—one can always alter the payoff of an unrealized state, changing Y^+ while leaving every historical return intact.

These features of the decomposition have immediate implications for asset demand analysis. Since Y^+ is unobservable, fundamental demand $\mathcal{D}^i$ cannot be recovered from portfolio data: different combinations of preferences and latent mappings are always observationally equivalent. And since Y^+ shifts whenever beliefs over payoffs are revised, asset demand functions are not structural with respect to standard perturbations that occur during regular market functioning.

What do these considerations imply for the estimation of asset demand functions without structural models? A common approach in the literature is to estimate *individual* asset demand curves through quasi-exogenous variation in asset supply. For this approach to work, supply shocks must generate *ceteris paribus* variation in a single asset price, holding all other prices and payoffs fixed.

Unfortunately, this identification assumption is generically inconsistent with the principle of no arbitrage and equilibrium price determination. Under minimal conditions, an increase in the supply of a given asset reduces the marginal cost of a unit payoff in a given state (i.e., the *state price*) in all states where the asset pays off. But by no arbitrage, this must lead to a change in the prices of all other assets that pay off in overlapping states. Since such payoff overlap is generic for essentially all asset markets, we prove that the identification of individual asset demand curves from individual supply shocks is generically infeasible. Most strikingly, we show that supply shocks generically imply state price changes that differ *directionally* from those required to estimate a fixed demand curve.

This leaves the possibility of jointly estimating the entire $J \times J$ *system* of demand slopes, where J is the number of assets, using multiple independent shocks to the price vector. This requires both at least J linearly independent price changes and that Y^+ remains fixed across all experiments. The first is a standard rank condition that arises in many settings; the second is the binding constraint. Since our decomposition shows that belief revisions generically shift all demand functions, it is implausible that the econometrician observes multiple independent shocks to the same demand system. Moreover, standard shocks used in the literature—such as central bank interventions or index inclusions—directly shift the demand system by altering future payoffs.

One might hope that imposing weak statistical structure on Y—for instance, through a factor model for asset returns—is enough to make progress. To investigate this, we use random matrix theory to study the statistical properties of the Moore-Penrose inverse Y^+ for factor-structured payoff matrices. Our results show the inverse mapping is generically ill-conditioned: the *sign* of any given element of Y^+ is a coin flip in large economies, and two economies sharing identical factor structures but different idiosyncratic payoff realizations have sign-independent inverses. Controlling for factor exposures therefore provides no systematic correction for the misalignment between supply shocks and the price variation required for demand estimation. We confirm these predictions numerically through Monte Carlo simulations and empirically using payoff data from S&P 500 stocks.

We summarize our results as a trilemma: given observational data, one cannot jointly maintain (i) no-arbitrage asset pricing, (ii) investor preferences over payoffs, and (iii) model-free identification of structural asset demand functions.

Our findings suggest a critical role for structural models in asset demand analysis. Since Y^+ cannot be identified from data, two models that agree on all observable implications of the data can imply arbitrarily different demand elasticities. For example, Fuchs, Fukuda, and Neuhann (2025) show that the logit model of Koijen and Yogo (2019) can infer an elasticity below one even if the true elasticity is infinite. Estimated asset demand elasticities should therefore not be treated as credible calibration targets, and should be evaluated on the plausibility and robustness of the assumed mapping rather than empirical fit.

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 that 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: investors form portfolios, marginal valuations depend on concurrent holdings of all other assets, the mapping from products to characteristics is latent, and choice is continuous. This fundamental non-separability of asset valuations under a latent mapping 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.

To overcome these challenges, structural models of asset demand must accurately account for the cross-asset linkages and price spillovers inherent to portfolio choice. Fuchs, Fukuda, and Neuhann (2025) show that the prominent logit approach in Koijen and Yogo (2019) can exhibit large biases in standard portfolio choice models with asymmetric substitution between assets. While our analysis in this paper focuses on contemporaneous cross-asset spillovers, similar 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. Perhaps most closely related to this paper is Haddad, He, Huebner, Kondor, and Loualiche (2025), who aim to recover relative demand elasticities from supply shocks without a structural model. Our findings suggest that their approach must impose theoretical restrictions if the estimated elasticities are to have a structural interpretation.

2 Framework

2.1 Environment

We study a canonical portfolio choice model. A mass of potentially heterogeneous investors I decide how much to consume at $t = 0$, and how to invest their savings to consume at $t = 1$. There are J financial assets, each of which yields a random payoff at date 1. Uncertainty is represented by a set of Z states of the world, one of which is realized at date 1. The payoff of asset j in state z is $y_j(z) \geq 0$. We denote by $Y \equiv (y_j(z))_{j,z}$ the $J \times Z$ matrix of cash flows. Since matrix Y reflects investors' beliefs about state-contingent payoffs, it is unobserved by the econometrician. We denote by $\pi \equiv (\pi_z)_z$, where $\pi_z \in (0, 1)$ is the probability of state z.

We are agnostic about the determinants of asset payoffs, and assume investors take the payoff matrix as given. However, in general the payoffs of a given asset are the sum of a direct cash component (i.e., dividends) and its expected resale price (i.e., the expected state-contingent market price). As we will discuss in more detail later, this means that one cannot easily assume that Y is a physical constant that remains fixed across time periods or economic regimes.

At time zero, each investor chooses a *portfolio* to maximize the expected utility of the state-contingent consumption across both dates. A portfolio is a vector of asset positions $a^i \equiv (a^i_j)_{j=1}^J \in \mathbb{R}^J$, where element a^i_j is the investor's holdings of asset j. Investor i's preferences are represented by a twice differentiable, strictly increasing and strictly concave von Neumann-Morgenstern utility function u^i.

Investors are competitive and take prices as given. The price of asset j is p_j, and time-zero consumption is the numeraire (or *outside asset*) with price normalized to one. Investor i is endowed with $e^i_j \geq 0$ units of asset j and $e^i_0 \geq 0$ units of the numeraire, and non-traded consumption endowments $w^i_0 \geq 0$ and $w^i(z) \geq 0$ at date 0 and in state z, respectively. Denote by $e^i \equiv (e^i_j)_j$ and $w^i \equiv (w^i(z))_z$. Portfolio choice may be curtailed by exogenous constraints: the investor must choose a portfolio from the set of feasible portfolios Φ^i, which we assume is a convex subset of $\mathbb{R}^J$.

Investor i's *portfolio choice problem* can then be formally stated as:

$$
\begin{aligned}
\sup_{a^i \in \Phi^i} \quad & (1-\delta^i)u^i(c^i_0) + \delta^i \pi \cdot u^i(c^i) \\
\text{s.t} \quad & c^i_0 = e^i_0 - p \cdot (a^i - e^i) + w^i_0 \quad \text{and} \\
& c^i = Y^{\mathrm{T}} a^i + w^i.
\end{aligned} \tag{PCP}
$$

Investor *i's asset span* $\mathcal{S}^i$ is set of payoff profiles that can be achieved through some feasible portfolio. That is,

$$
\mathcal{S}^i \equiv \{Y^{\mathrm{T}} a^i \in \mathbb{R}^Z \mid a^i \in \Phi^i\}. \tag{1}
$$

The portfolio choice problem embeds the canonical notion of *preferences over payoffs*: investors value state-contingent consumption and demand assets *instru-*

mentally for the payoffs they provide, not because they provide direct utility. Our results are robust to including a direct utility from holdings but are more sharply stated without them—all we require is that investors have at least some preferences over payoffs. A solution to this problem is jointly determined by several parameters: (i) the utility function u^i and rate of time preference δ^i, (ii) initial wealth w_0^i and state-contingent endowments w^i, which are demand shifters that shift state-specific marginal utility, (iii) portfolio constraints which determine the set of feasible portfolios Φ^i, and (iv) the payoff matrix Y and probability distribution π. We call the utility function, rate of time preference, demand shifters, and portfolio constraints *preference parameters*, which we denote by

$$\Theta^i \equiv \left(u^i, \delta^i, w_0^i, w^i, \Phi^i\right).$$

The optimal portfolio also depends on payoff matrix Y, which determines the mapping from asset positions to state-contingent payoffs, and the probability distribution π, which determines weights on states of the world. Since these objects do not pertain to investor preferences, we refer to these as *external parameters*.

Asset Demand Functions. A solution to problem (PCP) can be written in terms of J Marshallian *asset demand functions* which map parameters and the asset price vector into portfolio holdings. That is, the asset demand functional of investor i is

$$a^i(\cdot \mid \Theta^i, Y, \pi) : \mathbb{R}^J_{++} \to \mathbb{R}^J.$$

Standard portfolio choice theory shows that all asset demand functions generically depend on the entire vector of asset prices. That is, asset demand is *non-separable*.

In line with empirical practice, we will typically focus on identifying the $J \times J$ matrix of asset demand derivatives given a prevailing asset price vector p:

$$\mathcal{A}^i(\Theta^i, Y, \pi) \equiv -\frac{\partial a^i(p \mid \Theta^i, Y, \pi)}{\partial p^{\mathrm{T}}}.$$

This object characterizes asset demand within a neighborhood of price vector p.

In general, the econometrician observes neither preference parameters Θ^i nor external parameters (Y, π). The demand identification problem thus is to infer combinations of these parameters which determine asset-level demand functions and are invariant to perturbations or counterfactuals of interest.

2.2 Consistent Pricing Systems and No Arbitrage

Before analyzing the demand identification problem in more detail, we establish the importance of a consistent pricing system for all possible portfolios of assets. This motivates our use of no arbitrage to structure the pricing system.

Necessity of internally consistent prices. A defining feature of portfolio choice is that investors can flexibly bundle and unbundle assets to achieve desired payoff processes. For example, two assets with state-contingent payoffs $[1, 1]$ and $[1, 0]$ can be combined into a portfolio with payoff $[0, 1]$, or indeed *any* payoff in $\mathbb{R}^2$. Given continuous choice over assets, investors can thus choose among a *continuum* of potential payoff vectors whose mapping into portfolios depends on the unobserved payoff matrix Y. To permit inference about preferences from portfolio holdings, the econometrician must therefore impose a priori assumptions on the pricing system which can be used to construct prices for all feasible payoff vectors.

No arbitrage ensures consistent pricing. The canonical approach to ensuring consistent pricing in financial markets is the principle of *no arbitrage*, which asserts that pricing system should not permit trading strategies which offer "something for nothing." In particular, this principle states that there should not exist feasible trading strategies which offer strictly positive payoff at some date while offering weakly positive payoffs in all other states and dates.

Definition 1 (No Arbitrage) *There is no arbitrage if there does not exist a portfolio* $a^* \in \mathbb{R}^J$ *such that* $Y^{\mathrm{T}} a^* \geq 0$ *and either (i)* $p \cdot a^* \leq 0$ *and*$(Y^{\mathrm{T}} a^*)_z > 0$ *for some* z *or (ii)* $p \cdot a^* < 0$*.*

No arbitrage is a weak restriction which rules out the existence of profitable trading strategies that would be exploited by any investor with increasing preferences. Nevertheless, it is sufficient to ensure consistent pricing of *all* portfolios. In particular, the fundamental theorem of asset pricing shows that no arbitrage is equivalent to the existence of a vector of *state prices* $q \in \mathbb{R}^Z$ which serve as reference prices from which one can recover any asset price. State prices can be interpreted as the marginal cost of unit payoff in a given state of the world, so that asset prices are payoff-weighted sums of state prices. See Duffie (2001) for the proof.

Theorem 0 (Fundamental Theorem of Asset Pricing) *Let $\Phi^i = \mathbb{R}^J$. There is no arbitrage if and only if there exist state prices $q \in \mathbb{R}^Z_{++}$ such that asset prices satisfy*

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

Under no arbitrage, the prices of all traded portfolios are thus linear combinations of state prices, with weights determined by state-contingent payoffs. This yields a consistent pricing system that links asset prices to the marginal cost of what they offer to investors, namely state-contingent payoffs. Because of its foundational role in modern asset pricing, we adopt this pricing system as well.

Existence of Optimal Portfolios and Dimension Reduction. No arbitrage provides two additional advantages for asset demand analysis. First, it allows researchers to combine individual asset positions into aggregated portfolios while ensuring that the demand for and price of the bundle is internally consistent. Such portfolio aggregation is a foundational tool. For example, Koijen and Yogo (2019) aim to summarize asset demand using a small number of asset characteristics. Second, a canonical result—recapitulated below—shows that no arbitrage ensures the *existence* of a solution to the portfolio choice problem. Naturally, existence is a prerequisite for demand analysis in asset markets. See Duffie (2001) for the proof.

Proposition 0 (No arbitrage and the investor's problem) *Let $\Phi^i = \mathbb{R}^J$. Then there is a solution to (PCP) if and only if there is no arbitrage.*

Taken together, the principle of no arbitrage thus serves to ensure the internal consistency of asset demand systems while imposing only weak assumptions. While some trading frictions could prevent no arbitrage from holding exactly, as long as the frictions do not completely rule out general equilibrium price adjustments, our arguments hold.

Redundant assets. No arbitrage pricing is particularly salient in the presence of redundant assets (i.e., when there are multiple portfolios that deliver identical cash flow processes). In such cases, an arbitrarily small change in the price of a redundant asset immediately triggers an arbitrage opportunity with discontinuous changes in demand functions (see Example 3 in Appendix D.1 for an illustration). If such arbitrages do persist on the equilibrium path, it is infeasible to identify asset-specific demand functions (i.e., the slope of asset quantities with respect to variation in a single price) for redundant assets from observational data. For the remainder, we therefore focus on the case without redundant assets.

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

2.3 Outline of the Argument

Our argument proceeds in three steps. *First* (Section 3), we derive the decomposition $\mathcal{A}^i = (Y^+)^\mathrm{T}\mathcal{D}^i Y^+$, which separates asset demand slopes into fundamental preferences over payoffs and a latent mapping Y^+ from payoff demand into portfolios. Since Y^+ is unobservable in principle, asset demand functions are not structural with respect to the belief revisions that occur during ordinary market functioning.

Second (Section 4.1), we ask whether the latent mapping can be bypassed through supply shocks and show that it cannot: no-arbitrage forces prices of all payoff-overlapping assets to move jointly, so supply shocks generically produce price variation that is misaligned with the ceteris paribus requirements of demand identification.

Third (Section 4.3), we ask whether imposing factor structure is sufficient to make progress and show that it is not: using random matrix theory, we establish that Y^+ is generically ill-conditioned, with individual elements having the wrong sign with probability approaching one-half in large economies. Section 5 collects these findings into the trilemma and draws implications for the interpretation of estimated demand elasticities.

3 The Problem of the Unobservable Mapping

To understand basic properties of asset demand function, we begin by establishing a general decomposition of asset demand functions into two components: fundamental demand over state-contingent consumption, and a *latent mapping* which determines the asset portfolio required to achieve target state-contingent consumption profile. The first component reflects the standard notion of demand that is common to all demand analysis, in that it reflects preference parameters and willingness to pay for consumption in different states of the world. The latent mapping is unique to the case of financial assets, in that it reflects beliefs over future payoffs which guide how different assets must be combined with each other.

Our main result is that the latent mapping is fundamentally *unobservable*. This has two implications: (i) fundamental demand can never be identified from data on portfolio choices, and (ii) asset demand can be given a structural interpretation (that is, being invariant to perturbations) only if the generalized inverse of the payoff matrix Y remains fixed. Since payoff and forecast revisions occur across essentially all time horizons and financial markets, this suggests that asset demand is not a structural object in essentially all settings of interest.

3.1 Demand Decomposition and Non-Separability

To arrive at our decomposition, we must define an appropriate notion of demand functions for state-contingent consumption. We thus consider the following *con-*

sumption choice problem given a vector of state prices.

$$\begin{aligned} \max_{c^i \in \mathcal{C}^i} \quad & (1-\delta^i) u(c_0^i) + \delta^i \pi \cdot u(c^i) \\ \text{s.t} \quad & c_0^i + q \cdot (c^i - w^i) \leq w_0^i + q \cdot Y^{\mathrm{T}} e^i. \end{aligned} \tag{CCP}$$

A solution to problem (CCP) is a set of Z state-contingent consumption functions c^i which map the vector of state prices into a consumption profile. The budget constraint states that net purchases of consumption (over and above endowments) must be equal to the value of the investor's asset endowments. The set $\mathcal{C}^i$ encodes constraints on feasible consumption choices due to, e.g., incomplete markets or other portfolio constraints. By analogy with asset demand functions, we can therefore study the derivative of the Marshallian *consumption* demand curve with respect to *state prices*.

Definition 2 (Consumption demand) *The slope of consumption demand is a $Z \times Z$ matrix of Marshallian consumption demand slopes with respect to state prices,*

$$\mathcal{D}^i \equiv -\frac{\partial c^i}{\partial q^{\mathrm{T}}}$$

which depends on preference parameters, state probabilities π and state prices but is independent of the payoff matrix conditional on investor i's asset span $\mathcal{S}^i$ defined by (1).

Consumption demand functions thus have a certain robustness property with respect to small perturbations of the payoff matrix Y. When these perturbations do not affect the set of attainable payoffs, they do not alter consumption plans. As we will see, this is *not* the case for asset demand functions, which do depend directly on perturbations of Y.

However, consumption demand functions are not observable because neither state prices nor the payoff process are directly observed by the econometrician. However, they can be linked to observable portfolios and asset prices using the portfolio choice problem and no arbitrage pricing. This follows directly from the chain rule. Specifically, the consumption process induced by portfolio is $c^i = Y^\mathrm{T} a^i + w^i$. The portfolio yielding a target consumption process c^* thus is $a^*(c^*) = (Y^+)^\mathrm{T}(c^* - w^i)$, where Y^+ is the generalized Moore-Penrose inverse of Y. Differentiating asset demand with respect to asset prices therefore yields

$$\frac{\partial a^*(c^*)}{\partial p^\mathrm{T}} = (Y^+)^\mathrm{T} \frac{\partial c^*}{\partial p^\mathrm{T}} = (Y^+)^\mathrm{T} \frac{\partial c^*}{\partial q^\mathrm{T}} \frac{\partial q^\mathrm{T}}{\partial p^\mathrm{T}},$$

where the second equality follows from the chain rule. Next, observe that no arbitrage implies that asset prices p are related to state prices through $p = Yq$, and thus state prices can be related to asset prices through $q = Y^+ p$. Applying this observation to $\frac{\partial q^\mathrm{T}}{\partial p^\mathrm{T}}$ then yields the following decomposition.

Proposition 1 (Demand Decomposition) *If asset prices satisfy no arbitrage, then*

$$\mathcal{A}^i = (Y^+)^\mathrm{T} \mathcal{D}^i Y^+. \tag{3}$$

If preferences over state-contingent consumption are separable across states as in (PCP) *and there are no binding portfolio constraints, then one can further decompose*

$$\mathcal{A}^i = (Y^+)^\mathrm{T} \Pi^{-1} \tilde{\mathcal{D}}^i Y^+, \tag{4}$$

where Π is a diagonal matrix of state probabilities and $\tilde{\mathcal{D}}^i$ depends only on preferences.

The decomposition reveals that asset demand is generically *non-separable*: given a desired consumption process, the optimal position in any given asset is jointly determined by the payoff processes of *all* assets in the choice set. In particular, changing the payoff characteristics (state-contingent payoffs) of any given asset generically alters several elements of the inverse payoff matrix, triggering changes in the optimal demand for other assets.

A particularly stark illustration of this fact is that whether two assets are substitutes or complements depends on the attributes of other assets in the choice sets. To the best of our knowledge, this issue is distinct from other settings in industrial organization, which may consider flexible specifications in which two goods can be substitutes or complements, but these parameters are invariant to attributes of other goods. This however is a central feature of the portfolio choice.

We illustrate this with a simple 3 asset example, which satisfies the condition that any perturbations of the payoff structure do not change the asset span.

Example 1 (Complementarity and substitutability depend on other assets) *Consider a three-asset, three-state economy with the payoff matrix Y, where rows index assets:*

$$Y = \begin{bmatrix} 1 & \epsilon & 0 \\ \epsilon & 1 & 0 \\ \xi & \xi & 1 \end{bmatrix},$$

where $\xi, \epsilon \in (0,1)$. Thus, we have complete markets. Assuming homogeneous consumption elasticities, the cross-elasticity between Assets 1 and 2 is:

$$-\frac{\partial a_1}{\partial p_2} = \sum_z (Y^{-1})_{z,1}(Y^{-1})_{z,2} = \frac{\xi^2(1-\epsilon)^2 - 2\epsilon}{(1-\epsilon^2)^2}.$$

For a given ϵ we can find the value of ξ that sets the cross-elasticity to zero:

$$\xi^* = \frac{\sqrt{2\epsilon}}{1-\epsilon}.$$

Thus Assets 1 and 2 complements for $\xi < \xi^*$ and substitutes for $\xi > \xi^*$.

Figure 1 illustrates the change in complementarity as a function of ξ. To understand the intuition for this example, consider first a very low value of ξ. In this case, Asset 3 plays an insignificant role in the payoff of the first two states and Assets 1 and 2 are good hedges for each other in states 1 or 2. This force is stronger when ϵ is smaller. Next, consider a very large value of ξ: Asset 3 now replicates the payoffs of Assets 1 and 2 in states 1 and 2 ever more closely, crowding out the hedging roles of both. When the price of Asset 2 rises, investors turn to Asset 3 as an alternative hedge for state 2, but in doing so they simultaneously acquire exposure to state 1. This reduces their demand for Asset 1 as well. The two assets thus become substitutes not through any direct relationship between them, but because they share a common "competitor." This force is stronger when ϵ is larger.

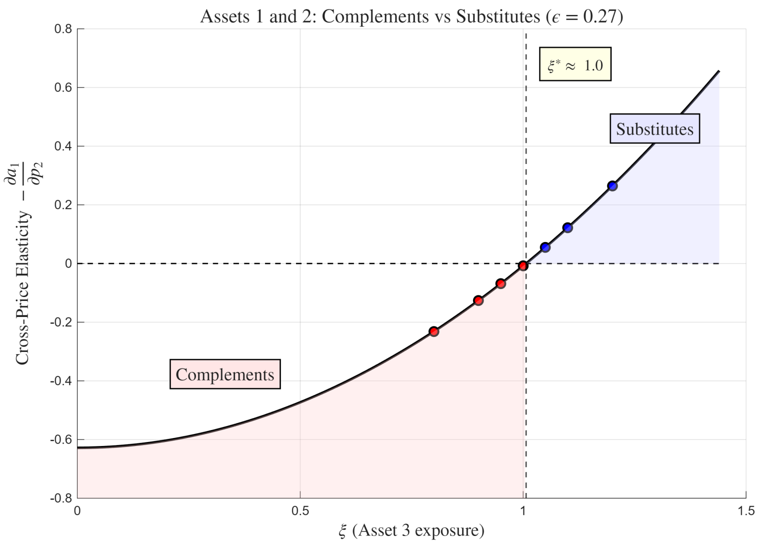


Figure 1: Switch in complementarity given ξ in Example 1

3.2 Unobservable Mapping

We have shown that asset demand functions—as well as the degree of complementarity between any two assets—is determined by global properties of the inverse payoff matrix Y^+. We next show a fundamental constraint on asset demand analysis: the mapping from fundamental preferences to portfolio choices is unobservable because the generalized inverse payoff matrix Y^+ cannot be identified from any finite sample of observed payoffs *even if the payoff matrix is assumed to be stable.*

Proposition 2 (Non-identification of the Latent Mapping Y^+) *Consider any finite sample of realized asset payoffs $\mathcal{S}$. Then there exist arbitrarily many candidate payoff matrices Y which are observationally equivalent given $\mathcal{S}$ but have different generalized inverses.*

Intuitively, the proposition holds because a researcher can always alter the payoff of a state of the world that has not yet been realized in the data, and altering payoffs in this state changes Y^+ but leaves all observed historical returns identical. This fundamental non-identification of Y^+ implies that asset demand estimation must reckon with two commingled identification problems: that of fundamental preferences (i.e., the standard identification problem that is common to all demand estimation exercises), and that of the latent mapping linking the primitive object of preferences (payoffs) to observed choices (asset positions).

An important implication of this is that fundamental demand can never be identified from observed portfolio choices. The reason is that observed choices and preferences pertain to different objects, namely assets and payoffs, and the mapping between the two is unobservable. As such, observed asset choices can always be rationalized by different combinations of preferences and the latent mapping.

Corollary 1 (Non-recoverability of $\mathcal{D}^i$) *For any $\mathcal{A}^i$, the consumption demand slope $\mathcal{D}^i$ is generically not uniquely determined absent knowledge of Y^+.*

Proof. The result follows because Y^+ cannot be identified from data (Proposition 2) and different inverses Y^+ generically induce different $\mathcal{D}^i$ through equation (3) even for the same $\mathcal{A}^i$. ■

3.3 Asset Demand is Not Structural under Weak Conditions

We now use our decomposition to provide conditions under which demand functions are *structural* in the sense of Hurwicz (1962) and Marschak (1953): invariant to relevant perturbations to the economic environment. To formalize this, we allow all model primitives to vary with a latent variable $\omega \in \Omega$ which we call the *economic environment*. We then say that a demand function is *structural* with respect to a class of perturbations if it is invariant across all ω within the class.

Definition 3 (Structural Demand) *Let $\mathcal{P} \subseteq \Omega \times \Omega$ be a class of perturbations, where each $(\omega, \omega') \in \mathcal{P}$ represents a transition from environment ω to ω'. We say that a demand function $\mathcal{F}$ is structural with respect to $\mathcal{P}$ if:*

$$\mathcal{F}(\omega) = \mathcal{F}(\omega') \quad \textit{for all } (\omega, \omega') \in \mathcal{P}.$$

The structural properties of demand thus depend on the perturbations under consideration. While these are often application specific, we emphasize perturbations that are of particular interest to financial markets: *unobserved* revisions to payoff expectations Y and state probabilities π which occur in the regular course of financial market operations. If asset demand is not structural with respect to these perturbations, there can be no guarantee that demand is structural with respect to other interventions or perturbations either.

Our decomposition then implies two main results: consumption and demand functions cannot be jointly be structural with respect to perturbations that vary the payoff matrix, and asset demand functions can be structural only if consumption preferences are a specific function of payoff parameters themselves. This contradicts the fundamental dichotomy between investor preferences and asset characteristics which permits demand analysis in the first place.

Proposition 3 (Non-structural demand) *Consider a class of perturbations with unobservable changes to Y or π. Then:*

1. *Asset demand function $\mathcal{A}^i$ and $\mathcal{D}^i$ cannot both be structural.*
2. *If asset demand functions are to be structural, then fundamental preferences must respond to any shock to future prices, dividends, or probabilities. In particular, we must have $\mathcal{D}^i = Y^T B^i Y$ for an arbitrary B^i, so that $\mathcal{A}^i = B^i$.*

That is, one cannot maintain the assumption that demand functions are structural unless one is also willing to make the assumption that probabilities and payoffs must remain fixed. This presents a sharp constraint on asset demand analysis because the assumption is (i) unverifiable, and (ii) it rules out that asset demand is structural with respect to interventions whose *goal* is to shift payoff expectations. This includes quantity-based policy experiments such as quantitative easing or foreign exchange interventions which aim to influence broader economic conditions. Such applications are of central interest to much of the literature.

4 The Problem of Identification

We now establish conditions under which asset demand can be identified from observational data on portfolio holdings and asset prices. We provide two main results. First, we show that the canonical approach to estimating demand curves—namely, exogenous shocks to the supply of a given asset—does not identify individual demand curves because they generically fail to produce ceteris paribus variation in a single asset price. Second, we ask whether multiple supply shocks can be combined to identify the $J \times J$ matrix of demand slopes $\mathcal{A}^i$. We show this to be possible if (i) the econometrician observes at least J linearly independent, exogenous shocks to the price vector and associated portfolio responses and (ii) the unverifiable assumption that the latent mapping Y^+ remains fixed across all such experiments. While the requirement of J shocks is a standard rank condition that also arises other settings with demand complementarities, the second is specific to financial markets, where the mapping between assets and characteristics is unobservable. Since forecast revisions are a defining function of financial market, this assumption is both strong and unverifiable in essentially all settings of interest. Furthermore, there is a natural tension: many experiments which shift prices today are likely to shift future prices as well, but this would also change Y^+, which contains resale prices.

4.1 Supply Shocks Produce Misaligned Price Variation

The canonical approach to estimating demand curves is to rely on supply shocks to provide suitably exogenous variation in a given price. With demand complementarities and long-lived assets, a central endogeneity concern is that the supply shock creates correlated changes in other asset prices (so-called *price spillovers*), thereby contaminating the price variation needed to identify a particular demand slope. We now show that this problem is generic under no arbitrage: except in the implausible knife-edge case where assets never pay off in overlapping states of the world, even perfectly exogenous supply shocks must always induce price spillovers.

Ideal experiment. We first define the price variation necessary to identify a particular asset demand slope. Given that asset demand exhibits demand complementarities and depends on expected future resale prices, we require that all other prices and all future payoffs must remain unchanged. To understand whether such variation is obtainable under even ideal conditions, we study the price changes induced by *perfectly exogenous supply shocks that leave all future payoffs unchanged.*

Under preferences over payoffs, it is useful to describe the ideal experiment in terms of state prices. The investor observes asset prices p and payoff matrix Y. Under no arbitrage, prevailing asset prices imply the state price vector

$$q = Y^{+} p, \tag{5}$$

where Y^{+} is the Moore-Penrose pseudo-inverse of Y.[1] A hypothetical *pure price shock* to asset j thus induces a specific state price change which is fully determined by Y^{+}.

Lemma 1 (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*

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

The assertion follows immediately from equation (5). Identifying asset demand thus requires shocks which generate the state price variation $\Delta \mathbf{q}_j^{\text{ideal}}$ associated with the ideal experiment.

Measurement using supply shocks. Since pure price shocks are rarely observed, we now study whether even perfectly exogenous supply shocks can generate the variation required by the ideal experiment. 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

[1] 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 exist multiple feasible state price vectors. As is standard, we select the minimum norm solution with pseudo-inverse $Y^{+} = Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}$, which also arises endogenously in our demand decomposition.

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 call this property *downward-sloping consumption demand*. Since our definition is written directly in terms of state prices, it must be understood purely in terms of fundamental preference parameters.

Definition 4 (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*

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

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 4 imposes no assumptions on V's off-diagonal entries, which capture potential *direct* preference-based spillovers across state prices in response to a supply shock. The existence of such spillovers depends on the economic model. The canonical model with additively separable utility over consumption (as in Section 2) has zero off-diagonal elements. Example 2 below illustrates this with a representative investor. Non-separable models such as recursive utility (Epstein and Zin, 1989; Kreps and Porteus, 1978) or more general aggregators instead generally imply non-zero off-diagonal elements. Since spillovers are the main threat to identification, the identification challenge is generically *weaker* when there are no preference-based spillovers in state prices. To provide favorable conditions for identification, we thus assume that no such spillovers exist.[2]

[2]The only case in which non-diagonal V can undo price spillovers occurs when the off-diagonal elements in V exactly offset the cross-asset restrictions implied by no arbitrage. However, V is determined by preferences and aggregate endowments while the no-arbitrage relation depends only on the payoff matrix Y. Hence there is no economic reason for such a mechanical offset to occur. More generally, Section 4.3 shows that, for large matrices, 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 meaningfully altering V.

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.*

Example 2 (V in an additive separable representative-agent model) *In a standard representative-agent model with additive separable preferences over consumption, state 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 pricing kernel is a strictly positive diagonal matrix,

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

Supply Shocks Do Not Generate the Ideal Experiment. We now show that supply shocks generically fail to produce the ideal experiment. We consider two definitions of alignment between supply shocks and the ideal experiment: (i) that induced state price changes are identical to those of the ideal experiment (up to a scalar multiple), and (ii) that the induced state price changes are of the same *sign*.

While the first condition is required to exactly identify a demand slope, the second captures the much weaker requirement that the supply shock should at least trigger *directionally consistent* changes in the cost of consumption. If this condition 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 lead to large biases when estimating substitution patterns.

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

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.*

We can then state our main result of this section, which is 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 that state. That is, in order to satisfy the minimal requirement that the induced state price variation is of the same sign as in the ideal experiment, there must be no assets with overlapping payoffs.

Definition 5 (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 (Supply Shocks Induce Misaligned Price Variation) *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.

The conditions set out 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 where overlap in payoffs (that is, concurrent non-zero dividends and/or resale values) is the norm, not the exception. It is therefore striking that, outside of these knife-edge restrictions, supply shocks do not even guarantee *directional* alignment with the ideal experiment. As such, supply shocks generically fail to identify structural asset-level demand slopes in essentially all settings of interest. In Appendix B, we also illustrate our findings using a simple example economy based on Fuchs, Fukuda, and Neuhann (2025).

Asset-by-asset misalignment. Theorem 1 shows that there must be misaligned price variation for at least one asset in the payoff menu (that is, at least one row of the payoff matrix). This allows the possibility that misalignment may not occur for *some* assets (although this cannot be verified without knowledge of the payoff matrix). The next proposition further strengthens our result by providing a weak condition for which misaligned price variation is guaranteed for *every asset*.

Proposition 4 *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 $(Y^+)_{z,j} < 0$.*

4.2 Identification from Multiple Shocks Requires Fixed Payoffs

So far we have shown that even perfectly exogenous supply shocks generically produce price variation that is contaminated by cross-asset spillovers. While this means that one cannot identify specific asset demand curves from asset-level supply shocks, one might yet jointly identify the entire matrix $\mathcal{A}^i$ by combining sufficiently many independent supply shocks.

We now show that this is the case only if (i) the econometrician has access to at least J exogenous supply shocks which provide linearly independent variation in the price vector—a standard rank condition—and (ii) that the unobservable mapping Y^+ remains fixed across all supply shocks. The second condition is a central challenge in financial markets, where the mapping from goods to characteristics is unobserved and subject to frequent revisions over essentially any time horizon.

Setting. We consider an idealized scenario in which the econometrician observes K independent shocks to the price vector and interprets investors' portfolio responses under a set of maintained assumptions $\mathcal{M}$. We refer to each price shock as an *experiment*, and assume that they are linearly independent. Since there are J assets, it is sufficient for our argument to consider the case where the number of experiments is weakly smaller than the number of assets: $K \leq J$. Let $\mathcal{O}_P$ denote the observed price changes, and $\mathcal{O}_{a^i}$ the observed portfolio response for investor i. We have the following standard definition of identification.

Definition 6 (Identified Demand Functions) *Two asset demand functions $a^i(\cdot \mid \Theta^i, Y, \pi)$ and $a^i(\cdot \mid \tilde{\Theta}^i, \tilde{Y}, \tilde{\pi})$ are observationally equivalent given data (O_P, O_{a^i}) and maintained* *assumptions $\mathcal{M}$ if both satisfy $\mathcal{M}$ and, for each observed price vector $p \in O_P$,*

$$a^i(p \mid \Theta^i, Y, \pi) = a^i(p \mid \tilde{\Theta}^i, \tilde{Y}, \tilde{\pi}) = O_{a^i}.$$

Demand slope $\mathcal{A}^i$ is identified under maintained assumptions $\mathcal{M}$ if all observationally equivalent demand functions imply the same slope:

$$\mathcal{A}^i(\Theta^i, Y, \pi) = \mathcal{A}^i(\tilde{\Theta}^i, \tilde{Y}, \tilde{\pi}).$$

We have already established that $\mathcal{D}^i$ is not identified because Y^+ is not observable. We now show that asset demand is identifiable only if the econometrician observes J experiments *and* maintains the unverifiable assumption that the unobservable mapping Y^+ is fixed across all experiments.

Proposition 5 (Necessary Conditions for Identification of $\mathcal{A}^i$) *Let (O_P, O_{a^i}) denote a data set of K observed price vectors and associated portfolio positions for investor i. Generically, $\mathcal{A}^i$ is identified only if the econometrician observes $K = J$ independent experiments and maintains the assumption that Y^+ is fixed across all experiments.*

Identification thus necessarily relies on the unverifiable assumption that the unobservable inverse payoff matrix is fixed across at least J shocks to the price vector. This finding relates our work to Haddad, He, Huebner, Kondor, and Loualiche (2025), who aim to recover an asset elasticity matrix without a fully specified structural model by combining (i) cross-sectional variation in asset-level holdings, and (ii) time series shocks to the prices of certain factor portfolios. Proposition 5 shows that this approach identifies a well-defined and stable elasticity matrix only if the unobserved payoff matrix remains fixed over time. However, changes in expected payoffs are a natural byproduct of financial market activity, including risk sharing, investment, or price discovery. (Section 4.3 shows that even approximate stability in the payoff matrix does not ensure a stable *inverse* payoff matrix, which is what matters for the stability of asset demand functions.)

Implications for Instrument Validity. Propostion 5 also sharpens the conditions for instrument validity in asset demand estimation: in addition to providing suitably exogenous variation in prices *today*, the instrument must *not* alter future expected payoffs. This rules out any shock to current prices that simultaneously induces changes in expected asset payouts or resale prices, such as central bank asset purchases or index inclusion events. The former are often used precisely to influence expectations over future market conditions, whereas index inclusion is known to alter return comovements and the level of prices.

4.3 The Latent Mapping is Ill-conditioned and Unstable

We have shown that asset demand functions are neither structural nor identifiable unless the latent inverse payoff matrix is assumed to remain fixed. These issues arise because one can at best observe a subset of *realized* payoffs, but not the matrix of *expected* payoffs, although it is the latter which matters for asset demand.

A potential solution to this problem is to use statistical information on realized returns to impose structure on the payoff process, and to hope that this structure is sufficient to ensure a stable mapping from preferences to asset holdings. The predominant approach in the literature is to impose a factor structure on payoffs, whereby asset returns are driven by a relatively small number of common factors. We therefore use random matrix theory to analyze the asymptotic properties of factor-structured payoff processes.

We find that the inverse payoff matrix is ill conditioned: the *sign* of any given element of the inverse payoff matrix is a coin flip. Hence even well-behaved factor structures yield poorly behaved latent mappings that can flip signs even with small changes to the payoff process. Monte Carlo simulations show that our theoretical limit results hold even for small J and Z. Appendix C shows that similar results hold in data from the S&P500. We conclude that imposing realistic statistical structure on the payoff matrix is *not* sufficient to ensure a well-behaved mapping from preferences to asset positions.

Random matrix approach. 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:

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

The analysis extends to multi-factor processes: see Remark 2 in Appendix A.4. As before, let Y^+ denote the Moore-Penrose pseudo-inverse of Y.[3] We impose the following assumptions.

Assumption 3 (Factor structure) *$(\alpha_j, \beta_j)_j$ are i.i.d., independent of $(f_z)_z$ and $(\varepsilon_{j,z})_{j,z}$, with finite second moments.*

Assumption 4 (Factor returns) *$(f_z - \overline{f})_z$ are i.i.d. with bounded, continuous, and symmetric densities around 0, and $\sigma_f^2 \equiv \mathbb{V}[f_z] < \infty$.*

Assumption 5 (Idiosyncratic shocks) *$(\varepsilon_{j,z})_{j,z}$ are i.i.d. across (j,z) with bounded, continuous, and symmetric densities around 0, and $\sigma_\varepsilon^2 \equiv \mathbb{V}[\varepsilon_{j,z}] > 0$. Factors and errors are mutually independent.*

Population objects and sequential limits. The properties of small random matrices are difficult to characterize with any generality. Theorem 2 thus considers the sequential limit $Z \to \infty$ followed by $J \to \infty$.[4] As $Z \to \infty$ the sample Gram matrix $G_Z \equiv \frac{1}{Z} YY^{\mathrm{T}}$ converges almost surely to the population second-moment matrix,

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

[3]The rank of Y equals J almost surely under Assumption 5, since the set of $J \times Z$ matrices with $\mathrm{rank}(Y) < J$ has measure zero. Hence $Y^+ = Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}$ a.s.

[4]The sequential limit $Z \to \infty$ followed by $J \to \infty$ is adopted for transparency of proof, not out of necessity. Our numerical simulations suggest that Z and J could be taken to infinity at the same time yet allowing for that would significantly complicate the proof.

The sign of $(Y^+)_{z,j}$ is asymptotically determined by the sign of the population quantity $(\Sigma^{-1}y_z)_j$, which can also be written as $v_j^\top \Sigma^{-1} y_z$ where $v_j \in \mathbb{R}^J$ is the j-th unit vector. We thus work directly with the following population objects:

1. *Individual sign probability.* For each fixed (j, z), define

$$\pi(j,z) \equiv \lim_{Z\to\infty} P\left((Y^+)_{z,j} > 0\right).$$

2. *Sign-agreement frequency.* Let Y and $\tilde{Y}$ be two payoff matrices generated by the same factor loadings $(\alpha_j, \beta_j)_j$ and factor realizations $(f_z)_z$ but independent idiosyncratic shocks $(\varepsilon_{j,z})$ and $(\tilde{\varepsilon}_{j,z})$. Define the sign agreement frequency

$$q(J,Z) \equiv \frac{1}{JZ}\sum_{j,z} \mathbf{1}\left(\text{sign}(Y^+)_{z,j} = \text{sign}(\tilde{Y}^+)_{z,j}\right),$$

where $\mathbf{1}(\cdot)$ denotes the indicator function, and its population limit

$$q(J) \equiv \operatorname*{plim}_{Z\to\infty} q(J,Z).$$

Theorem 2 characterizes the limits of these population objects as $J \to \infty$. We also use simulations to validate our results away from these limits.

Theorem 2 (Sign Instability of Y^+) *Under Assumptions 3–5, for almost every realization of $(\alpha_j, \beta_j)_j$, the following hold.*

(i) ***Individual coin flip.*** *For each fixed asset j and state z,*

$$\lim_{J\to\infty} \pi(j,z) = \lim_{J\to\infty} P\left((\Sigma^{-1}y_z)_j > 0\right) = \frac{1}{2}.$$

Moreover, the distribution of $(\Sigma^{-1}y_z)_j$ is continuous and centered at zero in the limit, so the positive and negative tails are mirror images of equal magnitude.

(ii) ***Factor structure knowledge is insufficient.*** *Let Y and $\tilde{Y}$ be two payoff matrices generated by the same factor loadings $(\alpha_j, \beta_j)_j$ and factor realizations $(f_z)_z$ but independent idiosyncratic shocks $(\varepsilon_{j,z})$ and $(\tilde{\varepsilon}_{j,z})$. The population sign-determining vari-*

ables $(\Sigma^{-1}y_z)_j$ *and* $(\Sigma^{-1}\tilde{y}_z)_j$ *are asymptotically independent for each fixed* (z, j)*, each with limiting sign probability* $\frac{1}{2}$*. Consequently,*

$$\lim_{J\to\infty} q(J) = \frac{1}{2}.$$

Knowledge of statistical properties of the return process is thus *not* sufficient to guarantee a well-behaved mapping from preferences to asset holdings. To the contrary, the latent mapping is generally ill-conditioned, with the sign of any given element being a coin flip. Hence the misalignment between supply shocks and the ideal experiment is pervasive for realistic payoff processes, and cannot be corrected for by controlling for factor exposures.

Calibration and Numerical Exploration Our theoretical results consider the limit $J, Z \to \infty$. We now study the behavior outside the limit using Monte Carlo simulations with payoff parameters that generate a share of idiosyncratic risk roughly consistent with the empirical data. Concretely, we assume:[5]

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

Figure 2 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. The left panel depicts the proportion of times any individual element of the matrix Y^+ is positive.[6] Thus, given $Y > 0$, almost half the elements of Y^+ have the wrong sign. The right panel compares the signs of $(Y^+)_{z,j}$ and $(\tilde{Y}^+)_{z,j}$ where both Y and $\tilde{Y}$ are generated from the same factor model and are thus indistinguishable in practice. We again observe that the signs coincide only 50% of the time, implying that there is no systematic way to correct for these sign errors.

[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]For each $Z \in \{150, 300\}$, we vary the number of assets $J \in \{2, 4, \ldots, 100\}$. We took the average of 1000 runs (of the Monte Carlo simulations). Figure 2 also depicts the 95% confidence interval.

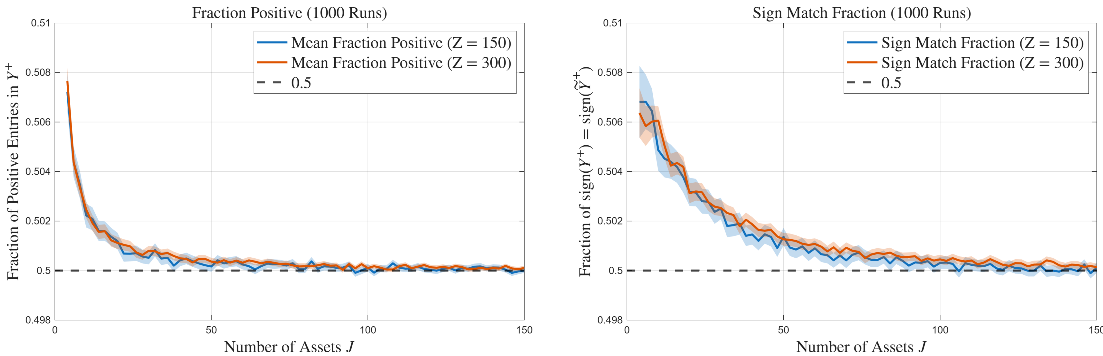


Figure 2: The Monte Carlo Simulation Results for Theorem 2. The left panel displays the empirical frequency of positive entries in Y^+, while the right panel shows the empirical frequency of sign matches between Y^+ and $\tilde{Y}^+$. Both panels report the results for $Z \in \{150, 300\}$ across various values of J, based on 1000 runs.

5 The Trilemma and its implications

We have established that the two principles of no arbitrage and preferences over payoffs sharply curtail the scope for non-parametric demand analysis in asset markets. We now summarize this result as a trilemma.

Theorem 3 (Trilemma) *Given observational data on portfolios, asset prices, and payoffs, one cannot jointly maintain (i) no-arbitrage asset pricing, (ii) investor preferences over payoffs, and (iii) model-free identification of structural asset demand functions.*

None of the stated conditions is easily discarded. No arbitrage is the prototypical internally-consistent pricing system, which ensures existence, consistency, and external validity of demand functions. Payoff-based asset valuation is the basic guiding principle of asset pricing. Since our definition of an asset is entirely generic, our results also apply equally to *portfolios* of primitive assets, which are themselves simply collections of payoffs. This leaves the assumption of constant payoffs. Unfortunately, this assumption is in tension with one of the basic functions of financial markets, which is price discovery—and thus revisions in expected payoffs—in response to news. It also cannot be directly verified in the data.

The trilemma is robust to small departures from its stated conditions. Relaxing no-arbitrage—for instance, by allowing for transaction costs or small arbitrage bands—replaces exact state-price equalities with inequalities but leaves the latent mapping Y^+ unobserved and ill-conditioned. Similarly, introducing non-pecuniary tastes or asset-level utility alters the fundamental demand object $\mathcal{D}^i$ but is orthogonal to the identification problem posed by Y^+: as long as investors retain any preference over state-contingent payoffs, the decomposition $A^i = (Y^+)^{\mathrm{T}}\mathcal{D}^i Y^+$ remains operative and Y^+ remains latent. The cross-asset linkages and sign instability documented in Theorem 2 depend only on the geometry of the payoff matrix, not on the preference specification. A model in which non-pecuniary considerations entirely dominate pecuniary ones is not, in any meaningful sense, a model of asset demand. Hence neither perturbation restores model-free identification of structural asset demand.

Our analysis suggests a critical role for structural models in asset demand estimation. However, since Y^+ cannot be identified from data, structurally estimated demand functions will depend sensitively on the assumed payoff structure. For example, Fuchs, Fukuda, and Neuhann (2025) show that the logit asset demand model proposed by Koijen and Yogo (2019) can yield low estimated demand elasticities because of the substitution patterns it assumes. The estimated demand elasticities should therefore be evaluated based on the validity and plausibility of the assumed payoff structure, not on empirical fit, which provides no information about whether the assumed latent mapping is correct. Our decomposition provides a framework for understanding what those restrictions imply and where misspecification is likely to be most consequential.

References

ALLEN, J., J. KASTL, AND M. WITTWER (2025): "Estimating Demand Systems with Bidding Data," Working Paper. [Cited on page 5.]

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

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 page 5.]

DUFFIE, D. (2001): Dynamic Asset Pricing Theory. Princeton University Press, third edn. [Cited on page 9.]

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 20.]

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 4.]

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 4.]

FUCHS, W., S. FUKUDA, AND D. NEUHANN (2025): "Demand-System Asset Pricing: Theoretical Foundations," Working Paper. [Cited on pages 3, 5, 22, 30, and 45.]

HADDAD, V., Z. HE, P. HUEBNER, P. KONDOR, AND E. LOUALICHE (2025): "Causal Inference for Asset Pricing," Working Paper. [Cited on pages 5 and 24.]

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 4.]

HE, Z., P. KONDOR, AND J. S. LI (2025): "Demand Elasticity in Dynamic Asset Pricing," Working Paper. [Cited on page 5.]

HURWICZ, L. (1962): "On the Structural Form of Interdependent Systems," in Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress, ed. by E. Nagel, P. Suppes, and A. Tarski, pp. 232–239. Stanford University Press. [Cited on page 16.]

IARIA, A., AND A. WANG (2020): "Identification and Estimation of Demand for Bundles," Working Paper. [Cited on page 4.]

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, 5, 9, and 30.]

KREPS, D. M., AND E. L. PORTEUS (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," Econometrica, 46(1), 185–200. [Cited on page 20.]

MARSCHAK, J. (1953): "Economic Measurements for Policy and Prediction," in Studies in Econometric Method, ed. by W. C. Hood, and T. C. Koopmans, no. 14 in Cowles Commission Monograph, pp. 1–26. John Wiley & Sons. [Cited on page 16.]

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.]

SHLEIFER, A. (1986): "Do Demand Curves for Stocks Slope Down?," Journal of Finance, 41(3), 579–590. [Cited on page 4.]

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 4.]

A Proofs of Propositions

A.1 Section 3

Proof of Proposition 1. First, the consumption process induced by portfolio a^i and payoff matrix Y is $c^i = Y^{\mathrm{T}} a^i + w^i$. Multiplying Y from the left, $Yc^i = YY^{\mathrm{T}} a^i + Yw^i$. Since YY^{T} is a $J \times J$ invertible matrix by Assumption 1, we have $a^i = (Y^+)^{\mathrm{T}} c^i - (Y^+)^{\mathrm{T}} w^i$. Since w^i does not depend on p, differentiating this expression yields $\frac{\partial a^i}{\partial p^{\mathrm{T}}} = (Y^+)^{\mathrm{T}} \frac{\partial c^i}{\partial p^{\mathrm{T}}}$. By no arbitrage, $p = Yq$ and hence $q = Y^+ p$. Then, equation (3) follows from the chain rule.

Second, under the stated conditions, the first-order condition with respect to c^i_z is:

$$\delta^i u'(c^i_z) = \lambda \pi_z^{-1} q_z,$$

where λ is the Lagrange multiplier on the budget constraint, which depends on the state price vector, as $\lambda = (1 - \delta^i) u'(c_0)$ where c_0 is optimal consumption at time 0. Differentiating this first-order condition with respect to $q_{z'}$ yields:

$$\delta^i u''(c^i_z) \frac{\partial c^i_z}{\partial q_{z'}} = \pi_z^{-1} \left(\frac{\partial \lambda}{\partial q_{z'}} q_z + \lambda \mathbf{1}(z = z') \right),$$

that is,

$$\frac{\partial c^i_z}{\partial q_{z'}} = \pi_z^{-1} \frac{1}{\delta^i u''(c^i_z)} \left(\frac{\partial \lambda}{\partial q_{z'}} q_z + \lambda \mathbf{1}(z = z') \right).$$

Letting $\tilde{\mathcal{D}}^i_{z,z'} \equiv \frac{1}{\delta^i u''(c^i_z)} \left(\frac{\partial \lambda}{\partial q_{z'}} q_z + \lambda \mathbf{1}(z = z') \right)$ and $\Pi \equiv \operatorname{diag}(\pi_1, \ldots, \pi_Z)$, we have:

$$\mathcal{D}^i = \Pi^{-1} \tilde{\mathcal{D}}^i.$$

Substituting this equation into (3) yields equation (4). ■

Proof of Proposition 2. Let $Y \in \mathbb{R}^{J \times Z}$ be a payoff matrix, and let $(z_1, \ldots, z_T)$ be any sample of realized states. Let $y \in \mathbb{R}^J$ be any vector with $y \notin \{y(1), \ldots, y(Z)\}$, where $y(z)$ is the z-th column of Y. Define $\tilde{Y} = [Y \mid y] \in \mathbb{R}^{J \times (Z+1)}$, where y corresponds to a state that does not realize in the sample.

With these in mind, first, we show that $\tilde{Y}$ is observationally equivalent to

Y: the realized return on every asset j in every period t is identical under Y and $\tilde{Y}$. To see this, since the additional state corresponding to y does not realize in the sample, the realized return on asset j in period t is $(Y)_{j,z_t}$ under both Y and $\tilde{Y}$ for all $j \in \{1, \dots, J\}$ and $t \in \{1, \dots, T\}$.

Second, the Moore-Pensrose inverses of the two matrices differ: $\tilde{Y}^+ \neq Y^+$. On the one hand, the Moore-Penrose inverse $Y^+ \in \mathbb{R}^{Z \times J}$ is given by

$$Y^+ = Y^\mathrm{T}(YY^\mathrm{T})^{-1}.$$

On the other hand, the Moore-Penrose inverse $\tilde{Y}^+ \in \mathbb{R}^{(Z+1) \times J}$ is given by

$$\tilde{Y}^+ = \begin{bmatrix} Y^\mathrm{T}(YY^\mathrm{T} + yy^\mathrm{T})^{-1} \\ y^\mathrm{T}(YY^\mathrm{T} + yy^\mathrm{T})^{-1} \end{bmatrix}.$$

Thus, in addition to $Y^+ \neq \tilde{Y}^+$, the $Z \times J$ block of $\tilde{Y}^+$ is different from Y^+ as $(YY^\mathrm{T} + yy^\mathrm{T})^{-1} \neq (YY^\mathrm{T})^{-1}$. ∎

Proof of Proposition 3. First, recall that $\mathcal{D}^i$ is the matrix of partial derivatives of optimal Marshallian consumption demand with respect to state prices, evaluated at given preferences and endowments. It is determined solely by the investor's preference ordering and budget set, and does not depend on Y directly. Hence, for any intervention that holds preferences and endowments fixed, $\mathcal{D}^i$ is unchanged.

In contrast, by Proposition 1, $\mathcal{A}^i = (Y^+)^\mathrm{T}\mathcal{D}^i Y^+$. Let $\tilde{Y}$ be a perturbation of Y, i.e., $\tilde{Y}$ is a $J \times \tilde{Z}$ non-negative matrix with $\text{rank}(Y) = J \leq \tilde{Z}$. Let $\tilde{\mathcal{A}}^i = (\tilde{Y}^+)^\mathrm{T}\mathcal{D}^i\tilde{Y}^+$. Then, $\tilde{\mathcal{A}}^i \neq \mathcal{A}^i$ with probability 1 (also, the set of $\tilde{Y}$ with $\tilde{\mathcal{A}}^i \neq \mathcal{A}^i$ is open and dense).

As Proposition 2 establishes that Y is not identified from realized return data, $\mathcal{A}^i$ cannot be identified as a structural object, and that $\mathcal{A}^i$ alone does not suffice to predict asset demand responses to interventions without additional identifying assumptions on Y^+. This proves the first part, and the second part is a contraposition of the first part. ∎

A.2 Section 4.1

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

$$Y^+ = DY^{\mathrm{T}}. \tag{6}$$

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 $z \in \{1, \ldots, Z\}$, it follows that

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

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

$$0 = \sum_{z=1}^{Z} y_j(z) y_k(z). \tag{7}$$

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 (7) implies that, for any $z \in \{1, \ldots, Z\}$, there exists at most one $j \in \{1, \ldots, J\}$ 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 1 (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 $z \in \{1, \ldots, Z\}$, there exists at least one $j \in \{1, \ldots, J\}$ such that $Y_{j,z} = y_j(z) > 0$, it can be shown that the diagonal matrix D in expression (6) is unique.

Proof of Proposition 4. 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:

$$\sum_{z=1}^{Z} y_k(z) Y^{+}_{z,k} = 1 \quad \text{for all} \quad k \in \{1, \ldots, J\}; \tag{8}$$

$$\sum_{z=1}^{Z} y_j(z) Y^{+}_{z,k} = 0 \quad \text{if} \quad j \neq k. \tag{9}$$

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 (9) 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 (8). For the sum to be strictly positive, there must exist at least one index z^* such that:

$$y_k(z^*) > 0 \quad \text{and} \quad Y^{+}_{z^*,k} > 0. \tag{10}$$

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 (10), 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.3 Section 4.2

Proof of Proposition 5. Suppose for contradiction that $\mathcal{A}^i$ is identified but Y^+ is not assumed fixed. Then, there exist two observationally equivalent parameter vectors (Θ^i, Y, π) and $(\tilde{\Theta}^i, \tilde{Y}, \tilde{\pi})$ with $Y^+ \neq \tilde{Y}^+$. By Proposition 2, such pairs exist for any finite dataset. By the demand decomposition of Proposition 1,

$$\mathcal{A}^i(\Theta^i, Y, \pi) = (Y^+)^T \mathcal{D}^i Y^+ \neq (\tilde{Y}^+)^T \mathcal{D}^i \tilde{Y}^+ = \mathcal{A}^i(\tilde{\Theta}^i, \tilde{Y}, \tilde{\pi})$$

generically, contradicting observational equivalence. Hence $\mathcal{M}$ must include the assumption that Y^+ is fixed. The requirement of $K = J$ independent experiments then follows from the fact that $\mathcal{A}^i$ is a $J \times J$ matrix, and point-identification of all its elements requires J linearly independent price vectors in O_P. ■

A.4 Section 4.3: Proof of Theorem 2

The proof proceeds in 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 $\frac{1}{Z} Y^{\mathrm{T}} \Sigma^{-1}$. Then, the population limits $\pi(j, z)$ and $q(J)$ equal expressions involving the population inverse Σ^{-1}, reducing both parts of the theorem to statements about $(\Sigma^{-1} y_z)_j$. 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., $(\nu_{z,j})_{z,j}$ in the main text). Then, we show the sense in which the deterministic shift is small compared to the stochastic component $(\nu_{z,j})_{z,j}$ when J is large by applying the Woodbury identity to Σ. With these in mind, the third step establishes part(i). The proof shows that the sign of the (z, j) element of Y^+, which is approximated by that of $\Sigma^{-1} Y^T$ (up to scaling) is asymptotically a fair coin flip. Finally, the fourth step establishes part (ii). The proof shows that the signs for two economies with a share factor structure but independent realizations of idiosyncratic shocks are asymptotically independent.

Hereafter, fix a realization of (α, β) for which all laws of large numbers used below hold. Probabilities are conditional on the loadings unless noted otherwise.

Step 1. In the first step, we replace the sample Gram matrix $G_Z \equiv \frac{1}{Z} Y Y^{\mathrm{T}}$ 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, where, as in the main text,

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

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 2 below 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, we decompose $(\Sigma^{-1} y_z)_j$ into the deterministic shift and the stochastic part symmetric around 0. Writing $y_z = \gamma + \beta(f_z - \bar{f}) + \varepsilon_z$ as in the main text, one can express

$$(\Sigma^{-1} y_z)_j = \underbrace{(\Sigma^{-1}\gamma)_j}_{\mu_j} + \underbrace{(f_z - \overline{f})(\Sigma^{-1}\beta)_j + (\Sigma^{-1}\varepsilon_z)_j}_{\nu_{z,j}}. \tag{11}$$

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

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

$$\begin{aligned} P((\Sigma^{-1} y_z)_j > 0) &= P(\mu_j + \nu_{z,j} > 0) \\ &= 1 - F_{\nu,j}(-\mu_j) = \frac{1}{2} + f_{\nu,j}(0)\mu_j + O(\mu_j^2), \end{aligned} \tag{12}$$

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

With these in mind, we now establish Lemma 2, 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 2 (Population replacement) *Fix J and (α, β). As $Z \to \infty$, the sample Gram matrix $G_Z \equiv \frac{1}{Z}YY^{\mathrm{T}}$ converges almost surely to Σ and hence $G_Z^{-1} \to \Sigma^{-1}$ a.s. Moreover,*

$$\lim_{Z\to\infty} \max_{1\leq j\leq J} \left| \frac{1}{Z}\sum_{z=1}^{Z} \mathbf{1}\left(\left(\tfrac{1}{Z}y_z^{\mathrm{T}}G_Z^{-1}\right)_j > 0 \right) - P\left((\Sigma^{-1}y_z)_j > 0 \right) \right| = 0 \quad a.s. \tag{13}$$

Consequently, conditional on (α, β),

$$\pi(j,z) = P\left((\Sigma^{-1}y_z)_j > 0 \right), \tag{14}$$

$$q(J) = \frac{1}{J}\sum_{j=1}^{J} P\left(\mathrm{sign}\big((\Sigma^{-1}y_z)_j\big) = \mathrm{sign}\big((\Sigma^{-1}\tilde{y}_z)_j\big) \right), \tag{15}$$

where y_z and $\tilde{y}_z$ denote the z-th columns of Y and $\tilde{Y}$.

Proof of Lemma 2. 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 $(\Sigma^{-1}y_z)_j = \mu_j + \nu_{z,j}$ has a continuous density at around 0 with value $f_{\nu,j}(0)$, we have:

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

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 (13). Then, the representations (14)–(15) follow immediately. ■

Step 2. The second step examines the structure of Σ^{-1}. The key insight is that $\Sigma = \sigma_\varepsilon^2 I_J + UU^\mathrm{T}$ is a rank-two perturbation of a scaled identity. Applying the Woodbury identity yields $\Sigma^{-1} = \sigma_\varepsilon^{-2} I_J - \sigma_\varepsilon^{-4} U A_J^{-1} U^\mathrm{T}$, where $A_J \equiv I_2 + \sigma_\varepsilon^{-2} U^\mathrm{T} U$ is a 2×2 matrix with $\|A_J^{-1}\| = O(J^{-1})$, because $U^\mathrm{T} U$ grows like J as more assets are added. The correction term $U A_J^{-1} U^\mathrm{T}$ is therefore small relative to the leading $\sigma_\varepsilon^{-2} I_J$ term. This has the following consequences established in the lemma:

Lemma 3 (Small deterministic shift) *Under Assumptions 3–5, the following hold uniformly in j as $J \to \infty$:*

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

Consequently, the stochastic fluctuation $\nu_{z,j} \equiv (f_z - \bar{f})(\Sigma^{-1}\beta)_j + (\Sigma^{-1}\varepsilon_z)_j$ is symmetric around zero with a continuous density and variance satisfying $\sigma_{\nu,j}^2 \to \sigma_\varepsilon^{-2}$ uniformly in j.

Proof. We apply the Woodbury identity to $\Sigma = \sigma_\varepsilon^2 I_J + UU^\mathrm{T}$:

$$\Sigma^{-1} = \sigma_\varepsilon^{-2}\left(I_J - U\left(I_2 + \sigma_\varepsilon^{-2} U^\mathrm{T} U\right)^{-1} \sigma_\varepsilon^{-2} U^\mathrm{T}\right). \tag{16}$$

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 $j \in \{1, \ldots, J\}$, 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 $\left(I_2 + \sigma_\varepsilon^{-2} U^\mathrm{T} U\right)^{-1}$ is of order $O(J^{-1}) = O(1) \cdot O(J^{-1}) \cdot O(1)$. Consequently, each entry in the matrix $U\left(I_2 + \sigma_\varepsilon^{-2} U^\mathrm{T} U\right)^{-1} \sigma_\varepsilon^{-2} U^\mathrm{T}$ 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.

Since we have established $(I_2 + \sigma_\varepsilon^{-2} U^{\mathrm{T}} U)^{-1} = O(J^{-1})$, substituting this back into Woodbury identity (16) and noting that $\gamma = U v_1$ 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 $U^{\mathrm{T}}\gamma = U^{\mathrm{T}} U v_1$, which is $O(J)$. Thus, the second term equals

$$\sigma_\varepsilon^{-4} U \big(I_2 + \sigma_\varepsilon^{-2} U^{\mathrm{T}} U\big)^{-1} U^{\mathrm{T}} U v_1 = \sigma_\varepsilon^{-2} U \cdot O(J^{-1}) \cdot O(J) = \sigma_\varepsilon^{-2} U \cdot O(1).$$

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 (16) 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

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

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

Lemma 3 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 the density at zero governing the linearization of the sign probability is determined primarily by the idiosyncratic variance σ_ε^2.

Step 3: Individual coin flip, proof of part (i). By Lemma 2, it suffices to analyze $P\left((\Sigma^{-1}y_z)_j > 0\right)$. By (12), we have:

$$\pi(j,z) = \frac{1}{2} + f_{v,j}(0)\mu_j + O(\mu_j^2). \tag{17}$$

By Lemma 3, $\mu_j \to 0$ uniformly as $J \to \infty$. Lemma 3 also implies that the density $f_{v,j}(0)$ bounded uniformly in j as $J \to \infty$. Thus, $\lim_{J\to\infty} \pi(j,z) = \frac{1}{2}$ for each fixed (j,z).[7] This establishes part (i).

Step 4: Sign instability, part (ii). By Lemma 2, $q(J)$ equals the average over j of

$$P\left(\text{sign}((\Sigma^{-1}y_z)_j) = \text{sign}((\Sigma^{-1}\tilde{y}_z)_j)\right).$$

Since Y and $\tilde{Y}$ share the same factor structure, both Gram matrices converge to the same Σ. Applying decomposition (11) to each matrix at a fixed (z,j) gives

$$\begin{aligned}
(\Sigma^{-1}y_z)_j &= \mu_j + (f_z - \bar{f})(\Sigma^{-1}\beta)_j + (\Sigma^{-1}\varepsilon_z)_j; \\
(\Sigma^{-1}\tilde{y}_z)_j &= \mu_j + (f_z - \bar{f})(\Sigma^{-1}\beta)_j + (\Sigma^{-1}\tilde{\varepsilon}_z)_j.
\end{aligned}$$

Both variables share the deterministic shift μ_j and the common factor component $(f_z - \bar{f})(\Sigma^{-1}\beta)_j$. By Lemma 3, $(\Sigma^{-1}\beta)_j = O(J^{-1})$, and thus the common factor component vanishes as $J \to \infty$. The residual stochastic terms $(\Sigma^{-1}\varepsilon_z)_j$ and $(\Sigma^{-1}\tilde{\varepsilon}_z)_j$ are independent (since $\varepsilon \perp \tilde{\varepsilon}$) and each symmetric around zero. In the limit, the two sign-determining variables reduce to independent symmetric random variables ξ and $\tilde{\xi}$. Hence,

$$\begin{aligned}
P\left(\text{sign}((\Sigma^{-1}y_z)_j) = \text{sign}((\Sigma^{-1}\tilde{y}_z)_j)\right) &\to P(\xi > 0, \tilde{\xi} > 0) + P(\xi < 0, \tilde{\xi} < 0) \\
&= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.
\end{aligned}$$

[7] An earlier working paper shows that $\frac{1}{J}\sum_{j=1}^{J} f_{v,j}(0)\mu_j \to \frac{1}{J} f_v(0)\Theta_1$, where the constant Θ_1 is determined by the model primitives.

Since this holds for each fixed j and the indicators are bounded, the dominated convergence theorem gives $\lim_{J\to\infty} q(J) = \frac{1}{2}$. ■

To conclude the proof of Theorem 2, two remarks are in order. Remark 2 shows that a multi-factor extension is possible. Remark 3 reiterates the economic interpretation of Theorem 2.

Remark 2 (Multi-factor extension) *For a K-factor model with*

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

the population Gram matrix takes the form

$$\Sigma = \sigma_\varepsilon^2 I_J + UU^{\mathrm{T}},$$

where

$$U \equiv \begin{bmatrix} \gamma & \sigma_{f,1}\beta^{(1)} & \dots & \sigma_{f,K}\beta^{(K)} \end{bmatrix} \in \mathbb{R}^{J\times(K+1)}.$$

Since $(I_{K+1} + \sigma_\varepsilon^{-2} U^{\mathrm{T}} U)^{-1} = O(J^{-1})$ for any fixed K, Lemma 3 carries over verbatim: both $(\Sigma^{-1}\gamma)_j$ and each $(\Sigma^{-1}\beta^{(k)})_j$ are $O(J^{-1})$. Hence both parts of Theorem 2 hold unchanged for any finite number of factors.

Remark 3 (Economic interpretation) *Theorem 2 sharpens the identification impossibility established in Theorem 1 of the paper. Theorem 1 shows that supply shocks generically fail to generate the correct direction of state-price changes for at least one asset. Part (i) here quantifies how severe this directional failure is for each state-asset pair individually: the sign of the required correction is a coin flip in large economies. Part (ii) shows that this instability cannot be resolved by obtaining a second sample from an economy with the same factor structure: even sharing the same systematic risk, two economies require corrections of opposite sign approximately half the time. This rules out any procedure for controlling directional errors that relies solely on the factor structure of payoffs.*

B Illustration in a General Equilibrium Model

We illustrate our findings using a simple example economy based on Fuchs, Fukuda, and Neuhann (2025). The decision problem is as in (PCP). For tractability, we assume that all investors are symmetric, face no portfolio constraints and have log utility.

Payoffs. There are two assets and two aggregate states of the world, both denoted by g (green) and r (red). The probability of state $z \in \{g, r\}$ is $\pi_z \in (0, 1)$. Table 1 depicts the payoff matrix. Parameter $\epsilon \in (0, 1)$ determines the complementarity between green and red assets. As $\epsilon \to 0$, green and red assets are perfect substitutes. As $\epsilon \to 1$, 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 1: Payoff matrix.

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 which we use to create price variation.

Asset demand. We will be interested in analyzing asset demand functions in a neighborhood around $s_g = 0$. We derive the demand functions a_g and a_r directly from the representative agent's portfolio choice problem:

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

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

$$(1-\delta)\frac{p_g}{E_0 - p_g(a_g - E_g) - p_r(a_r - E_r)} = \delta\pi_g \frac{1+\epsilon}{(1+\epsilon)a_g + (1-\epsilon)a_r} + \delta\pi_r \frac{1-\epsilon}{(1-\epsilon)a_g + (1+\epsilon)a_r}; \tag{18}$$

$$(1-\delta)\frac{p_r}{E_0 - p_g(a_g - E_g) - p_r(a_r - E_r)} = \delta\pi_g \frac{1-\epsilon}{(1+\epsilon)a_g + (1-\epsilon)a_r} + \delta\pi_r \frac{1+\epsilon}{(1-\epsilon)a_g + (1+\epsilon)a_r}. \tag{19}$$

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

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

$$a_g(p_g, p_r) = \delta \frac{(1 + p_g(1+s_g) + p_r)\left((1-\epsilon^2)p_g - ((1+\epsilon)^2 - 4\epsilon\rho)p_r\right)}{(p_g - p_r)^2 - (p_g + p_r)^2\epsilon^2}. \tag{20}$$

Varying only the green assets' price yields the standard own-price elasticity:

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

Misalignment between ideal experiment and supply shock. In the ideal experiment, the investor faces an exogenous increase in the price of the green asset p_g while p_r remains fixed. By Lemma 1, the induced change in state prices is

$$\Delta \mathbf{q}_g^{\text{ideal}} = \frac{\partial}{\partial p(g)}\begin{bmatrix} q(g) \\ q(r) \end{bmatrix} = \frac{1}{2\epsilon}\begin{bmatrix} 1+\epsilon \\ -(1-\epsilon) \end{bmatrix}. \tag{21}$$

A pure shock to $p(g)r$ thus *raises* the state price in state g, but *lowers* it in state r. This decrease in $q(r)$ is necessary to keep $p(r)$ unchanged.

Now consider how a supply shock s_g affects equilibrium state prices. Given resource constraints $c(z) = y_g(z)(1+s_g) + y_r(z)$, equilibrium state prices are

$$q(g) = \pi_g \frac{\delta}{1-\delta} \cdot \frac{1}{1+\frac{1+\epsilon}{2}s_g} \qquad \text{and} \qquad q(r) = \pi_r \frac{\delta}{1-\delta} \cdot \frac{1}{1+\frac{1-\epsilon}{2}s_g}. \tag{22}$$

Differentiating yields

$$\Delta \mathbf{q}_g^{\text{supply}} = \frac{\partial}{\partial s_g}\begin{bmatrix} q(g) \\ q(r) \end{bmatrix} = -\frac{1-\delta}{2\delta}\begin{bmatrix} (1+\epsilon)\cdot \frac{q(g)^2}{\pi_g} \\ (1-\epsilon)\cdot \frac{q(r)^2}{\pi_r} \end{bmatrix} < 0. \tag{23}$$

In contrast to the ideal experiment, a positive supply shock to the green asset thus decreases *both* state prices whenever $\epsilon < 1$. The simple reason is that the green asset pays off in both states of the world. As such, the supply shock generates a state price change Δq_g that is of the *wrong sign* compared to the ideal experiment. The only exception is when both assets are Arrow securities ($\epsilon = 1$).

Implications for demand elasticties. The misalignment between supply shock and ideal experiment can sharply bias observed behavior. In the ideal experiment, the investor is more willing to substitute away from green assets because the price of the red asset is unchanged. In the supply shock, substitution is tempered because the red asset is endogenously repriced. The resulting "elasticity" measure $\mathcal{E}_g^{\text{supply}}$ thus has an additional term which accounts the spillover to p_r:

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

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

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

We plot both measures in Figure 3. The two differ by order of magnitude for small ϵ. In this range, the two assets are close substitutes. In the ideal experiment

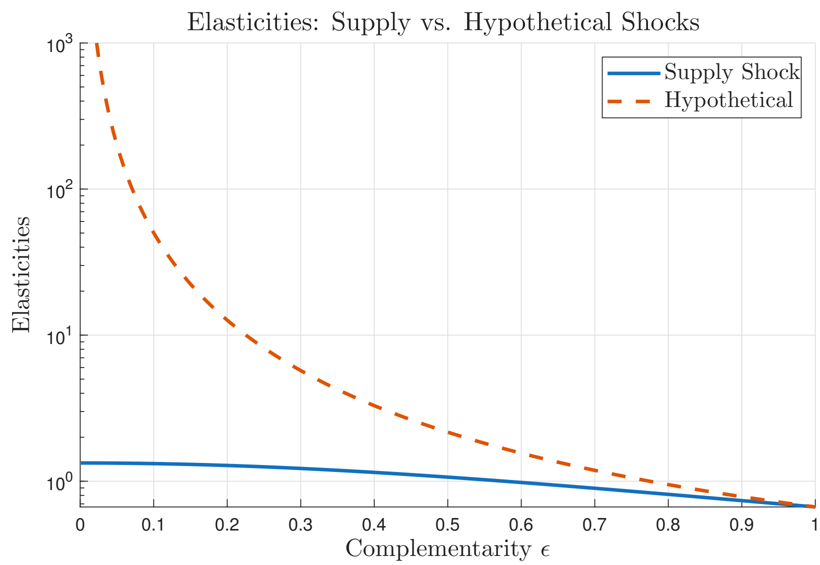


Figure 3: Ideal vs. supply-shock elasticities as a function of ϵ for $\delta = 2/3$ and $\pi_r = \pi_g = 1/2$. The ideal elasticity (solid line) diverges as $\epsilon \to 0$, while the supply-shock elasticity (dashed line) remains bounded. Both elasticities converge to $1 - \delta\pi_g = 2/3$ at the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

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. The only exception is when $\epsilon \to 1$ and the assets approach Arrow securities. In this case, there is no spillover across assets and thus no difference between the ideal experiment and the supply shock.

C Empirical Illustration

To further gauge the empirical relevance of our results, we conduct a simple empirical exercise in which we use payoff data from the *S&P* 500 to assess the alignment between (subsets of) the payoff matrix and its inverse. The exercise is not intended to be exhaustive, but simply illustrates the immediate relevance of the issues we discuss. The sample consists of 428 stocks that remained in the *S&P* 500 from 2020 to 2024. Since the true payoff matrix is latent, we construct (subsets) of it by sampling realized payoffs.

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 repeat this exercise ten times with replacement and report averages across all ten repetitions. Table 2 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. Taken together, 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 2: Results of our empirical exercise averaged over 10 iterations.

D Online Appendix

The Online Appendix is structured as follows. Appendix D.1 presents an example where redundant assets cause discontinuous demand. Appendix D.2 supplements Section 3 (specifically, Proposition 3). Appendix D.3 complements Section 4.1 by providing conditions under which Y^+ has the wrong sign for each state (Proposition 6), analogous to the asset-specific conditions in Proposition 4.

D.1 Section 2.2

We present an example in which an asset demand function exhibits discontinuity in the presence of redundant assets.

Example 3 (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.

1. *Suppose $\Phi^i = \mathbb{R}^3$. 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 (PCP) is no longer well-defined, and well-defined demand functions no longer exist.*

2. *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 $a^i_3 = -\chi$. *This can trigger discontinuities in optimal demand.*

D.2 Section 3

Remark 4 (Proposition 3) *Supposing that* $\mathcal{D}^i$ *is invertible, if* $\text{row}(Y) \neq \text{row}(\tilde{Y})$ *then* $\mathcal{A}^i \neq \tilde{\mathcal{A}}^i$.

To see this, for ease of notation, we introduce matrices $M_Y \equiv (Y^+)^T \mathcal{D}^i Y^+$ *and* $M_{\tilde{Y}} \equiv (\tilde{Y}^+)^T \mathcal{D}^i \tilde{Y}^+$. *We have* $\mathcal{A}^i \neq \tilde{\mathcal{A}}^i$ *if (and only if)* $M_Y \neq M_{\tilde{Y}}$.

Since $(YY^{\mathrm{T}})^{-1}$ *is an invertible* $J \times J$ *matrix, the range of* $(Y^+)^{\mathrm{T}}$ *satisfies:*

$$\text{Range}((Y^+)^{\mathrm{T}}) = \text{Range}(Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}) = \text{Range}(Y^{\mathrm{T}}) = \text{row}(Y).$$

Then, we consider the full product M_Y. *Since* $\mathcal{D}^i$ *is assumed to be invertible and* Y^+ *has full rank* J, *the product* $\mathcal{D}^i Y^+$ *is a* $J \times Z$ *matrix with rank* J. *Thus, invoking the properties of the range of a matrix product,*

$$\text{Range}(M_Y) = \text{Range}((Y^+)^{\mathrm{T}}) = \text{row}(Y).$$

Now, suppose toward a contradiction that $M_Y = M_{\tilde{Y}}$. *If two matrices are equal, then:* $\text{Range}(M_Y) = \text{Range}(M_{\tilde{Y}})$. *Substituting our previous result:*

$$\text{row}(Y) = \text{row}(\tilde{Y}).$$

This contradicts the initial hypothesis that $\text{row}(Y) \neq \text{row}(\tilde{Y})$. *Hence,* $M_Y \neq M_{\tilde{Y}}$, *and consequently* $\mathcal{A}^i \neq \tilde{\mathcal{A}}^i$. *The proof also states if* $\text{row}(Y^+) \neq \text{row}(\tilde{Y}^+)$ *then* $\mathcal{A}^i \neq \tilde{\mathcal{A}}^i$.

D.3 Section 4.1

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

Proposition 6 *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 6, 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 6. 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). ■

Demand-System Asset Pricing: Theoretical Foundations*

William Fuchs[†] Satoshi Fukuda[‡] Daniel Neuhann[§]

April 16, 2025

Abstract

Recent approaches to asset pricing use structural methods to estimate investor-level demand functions for financial assets. We show that cross-asset complementarities and price spillovers can significantly bias these estimates: if close substitutes exist, measured elasticities are near one even if true elasticities are near infinite. This reconciles low demand-system elasticities with higher theoretical benchmarks. Biases are smaller for less substitutable assets, such as broad portfolios or asset classes. Control variables lead to estimates of *residual* demand elasticities which may offer limited information about asset-level demand. We caution against interpreting estimated demand elasticities as structural parameters which remain stable under counterfactuals.

*Special thanks to Andres Almazan, Zhiguo He, Marco Loseto, Christian Opp (discussant), Andrey Ordin, and Andy Skrypacz for detailed conversations. We also thank Manuel Adelino, Hengjie Ai (discussant), Aydogan Alti, Philippe van der Beck (discussant), Alon Brav, Nicolae Garleanu (discussant), Valentin Haddad, Lars Hansen, Ben Hebert, Ali Hortacsu, Amy Huber, Paul Huebner, Travis Johnson, Peter Kondor, Jian Li, Erik Loualiche, Deborah Lucas, Aaron Pancost, Stavros Panageas (discussant), Anna Pavlova, Lukas Schmid, Dejanir Silva (discussant), Yannis Stamatopoulos, Sheridan Titman, Jonathan Wallen, Milena Wittwer, and seminar attendees at ASU Sonoran Winter Finance Conference, Bocconi, Cornell, Duke/UNC Asset Pricing Conference, 2024 Finance Theory Group Spring Meeting, 2024 FTG Conference on Bridging Theory and Empirical Research, 2024 FIRS conference, Finance Theory Webinar, Stanford GSB, UT Austin and Wharton for useful feedback. We thank Kevin Mei for excellent research assistance. Fuchs gratefully acknowledges the support from Grant PGC2018-096159-B-I00 financed by MCIN-AEI-10.13039-501100011033, and Comunidad de Madrid (Spain) through grants EPUC3M11 (V PRICIT) and H2019-HUM-5891.

[†]University of Texas at Austin, UC3M, FTG, and CEPR. william.fuchs@mccombs.utexas.edu.

[‡]Bocconi University and IGIER. satoshi.fukuda@unibocconi.it.

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

1 Introduction

Recent approaches to asset pricing following Koijen and Yogo (2019) involve the structural estimation of investor-level demand functions for financial assets. Advocates of this approach argue that detailed data on portfolio holdings can be used to structurally identify investor-level demand parameters for specific assets, and that granular descriptions of individual demand functions offer new insights into the functioning of financial markets, including the equilibrium response to a wide array of counterfactuals, such as shocks to the wealth distributions, investor preferences, or policy interventions. Perhaps the most striking claim in this literature is that demand elasticities for financial assets are orders of magnitude lower than in standard models (Koijen and Yogo, 2021).

While compelling in its motivation, the promise of the demand-system approach ultimately hinges on its ability to accurately identify structural parameters of interest. Yet existing demand estimation techniques, including the logit approach in Koijen and Yogo (2019), were originally developed for settings that differ substantially from portfolio choice and asset pricing. For example, the prototypical demand estimation in industrial organization considers market-level analyses of discrete choices over consumption goods. Such settings generally do not feature many considerations which are central to asset pricing, including cross-asset demand complementarities within portfolios, variable quantities, dynamic trading, and general equilibrium price determination. Hence it is an open question whether current demand approaches can indeed identify structural parameters in settings where these features are critical. We answer this question in a canonical asset pricing model (Lucas, 1978) enriched with heterogeneous tastes for financial assets.[1]

Our main finding is that current asset demand systems do not adequately account for cross-asset complementarities in portfolio choice, whereby the marginal value of an asset depends, in an asset-specific manner, on the investor's holdings of other assets. As such, asset demand systems may yield low *measured* demand elasticities even when true elasticities are near infinite. This offers a simple explanation for the striking difference between the low demand elasticities documented by leading demand systems estimates and the high demand elasticities obtained in standard models. More constructively, we

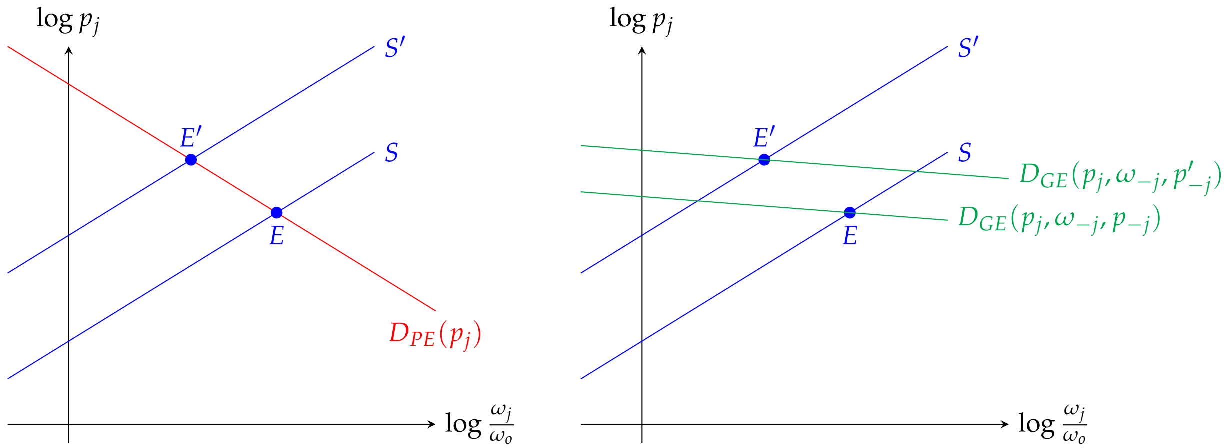
[1] Taste differences are critical because they generate cross-sectional heterogeneity in portfolios, as in the data. They also make it possible to construct demand shocks to some investors that are suitably orthogonal to the demand of other investors. This is required for identification of demand functions.

also show that this bias is smaller when assets are not particularly substitutable, as may be the case for highly aggregated asset classes with little overlap in the cash-flow distribution. Nevertheless, we caution that demand elasticities, even if well-measured, can be interpreted as structural parameters only under stringent additional assumptions.

A simple thought experiment is instructive. Consider an investor who must choose between three assets: two closely substitutable "inside assets," say Microsoft and Apple stocks, and a less substitutable "outside asset," say a bond. Now consider a supply shock to Microsoft which raises its price. All else equal, the investor would like to buy Apple and finance this trade by selling Microsoft (a *demand complementarity*). Given this substitution within inside assets, she then finds it optimal to leave her holdings of the outside asset roughly unchanged (*heterogeneous substitution*). Yet if many investors pursue the same strategy, the price of Apple must increase (a *price spillover*), taking away any individual investor's incentive to switch from Microsoft to Apple. In equilibrium, the investor's portfolio is thus relatively unresponsive to exogenous variation in the price of Microsoft even though she would have responded very rapidly had the price of Apple remained fixed. That is, demand complementarities and price spillovers create a disconnect between observed elasticities (which incorporate all equilibrium adjustments) and structural elasticities (which counterfactually presume that other asset prices remain fixed).

Given this disconnect, identifying structural elasticities from observational data requires appropriately accounting for heterogeneous substitution and price spillovers. Yet current approaches place stark restrictions on substitution patterns and price spillovers. For example, logit demand systems following Koijen and Yogo (2019) assume that complementarities and spillovers can be accounted for by measuring demand for inside assets *relative* to the outside asset. Given this restriction, price spillovers between inside assets are immaterial, and observed changes in relative portfolio shares identify the elasticity of relative demand. Naturally, biases arise when this restriction fails, as is the case when inside assets are more substitutable with each other than with the outside asset.

Figure 1 illustrates the nature of this bias. In both panels, we consider an exogenous supply shock from S to S', and identical observed equilibrium prices and quantities E and E'. The left panel depicts a demand system following the approach in Koijen and Yogo (2019) in which the demand curve determining the *relative* portfolio share of asset j is invariant in the quantities or prices of other inside assets. Given this assumption,



(a) Logit demand

(b) Demand with cross-asset complementarities

Figure 1: Elasticity measurement based on different demand system specifications. The left panel corresponds to logit demand for financial assets (Koijen and Yogo, 2019), whereby relative demand for a given asset is invariant in the prices and quantities of other assets (as in Partial Equilibrium). The right panel allows for cross-asset complementarities (as in General Equilibrium). The supply curves and the observed equilibrium allocations are identical in both panels. We use ω_j to refer to portfolio share of asset j, and ω_{-j} to denote the vector of portfolio shares of assets other than j. Analogous definitions hold for asset prices p_j. Demand is measured in units of portfolio shares relative to ω_o, the portfolio share of the outside asset.

observed portfolio changes are interpreted as a move *along* a relatively *inelastic* demand curve. The right panel depicts a demand system in which demand for asset j depends on the holdings and prices of other inside assets, and these respond endogenously to the supply shock. The observed demand response is now rationalized by *high* elasticities and a *shift* of the demand curve. Because standard models naturally generate demand curves with complementarities (even when measured in relative terms), this mechanism can account for the dramatic difference between low measured elasticities in the demand-system approach and much higher elasticities in standard models.

We derive this bias formally by decomposing the difference between measured and structural elasticities into the product of *demand complementarities* (i.e., the cross-elasticity between the focal asset and all potential substitutes and complements) and *price spillovers* (the equilibrium response of other asset to a shock to a given asset). Hence the bias is large whenever close substitutes are available and substitute assets are quickly repriced in response to shocks. That is, the bias is large precisely when markets are elastic.

One suggested approach to the problem of heterogeneous substitution and spillovers is the use of control variables. A given supply shock may trigger spillovers primarily to other assets with similar factor exposures. In this case, controlling for common factor exposures mitigates the resulting asymmetry in substitution. However, controls also directly alter the degree of substitutability between choices: two assets may be substitutable precisely *because* they have common exposures. The use of controls then yields demand elasticities defined over the *residual* cash flows unaccounted for by controls. If these residual cash flows are less substitutable than the asset itself, the resulting demand elasticities are naturally lower and may carry little information about asset-level elasticities.

In the final part of our analysis, we ask whether financial demand elasticities, even if well-measured, should be interpreted as structural parameters that are likely to be invariant under counterfactuals. To do so, we incorporate another feature that distinguishes financial markets from many goods markets, which is that assets are investment goods whose current value critically depends on their resale price. Using a dynamic variant of our model, we derive demand functions for financial assets that depend both on the investor's private tastes and her expectations of market returns. As in a beauty contest (Keynes, 1936), demand elasticities alone thus cannot distinguish whether an investors' demand is due to her own tastes or her expectations of others' tastes. Yet estimating counterfactuals in many cases requires assigning tastes to a particular investor.

We establish a related result for unobservable portfolio constraints. For a range of parameters, tastes for a given asset are observationally equivalent to unobserved mandates that constrain an investor's portfolio choice. Yet an unconstrained investor will respond differently to a counterfactual price shock than a constrained investor.

Related Literature

Demand-system asset pricing is grounded in structural estimation of investor-level portfolio choice functions. This is a sharp break from neoclassical asset pricing, which has little interest in asset quantities and instead focuses on price data disciplined by no arbitrage (Ross, 2004). It also differs from existing approaches that do emphasize quantities, such as classical theories of portfolio balance (Tobin, 1969), convenience yields in Treasury markets (Krishnamurthy and Vissing-Jorgensen, 2012), intermediary asset pricing (He and Krishnamurthy, 2013; Adrian, Etula, and Muir, 2014), capital flows due to index inclusion or other market frictions (Shleifer, 1986; Harris and Gurel, 1986), which emphasize *aggregate* demand effects in certain asset classes or markets, but stop short of structurally estimating investor-level demand functions for specific assets.[2] The fact that estimated demand systems appear to reveal that financial institutions exhibit low demand elasticities has also been used to argue that financial markets as a whole are inelastic, with implications for the equity premium (Gabaix and Koijen, 2020).

Demand elasticities implied by the aforementioned literature on capital flows are broadly similar to those found in demand-based approaches. However, the goal of this literature is not to isolate structural elasticities (in which all other prices are held fixed) from general equilibrium elasticities (which incorporate all adjustments). That both approaches find similar elasticities can therefore be explained by the fact that they ultimately estimate similar objects (namely, the general equilibrium elasticity). Our analysis here focuses on approaches that aim to estimate structural elasticities and parameters.

Several papers build on the logit demand system to study substantive questions, including effects the of counterfactual wealth distributions (Koijen, Richmond, and Yogo, 2024), global imbalances and currencies (Jiang, Richmond, and Zhang, 2023), corporate bond markets (Bretscher, Schmid, Sen, and Sharma, 2022; Darmouni, Siani, and Xiao, 2023), asset purchase programs (Breckenfelder and De Falco, 2023), bond market substitution (Nenova, 2025), and stock market competitiveness (Haddad, Huebner, and Loualiche, 2025). Davis, Kargar, and Li (2025) share our interest in accounting for low measured elasticities but use a partial equilibrium approach without endogenous price spillovers. While these studies extend the scope of asset demand systems in important ways, they do not address the specific issues of complementarities and spillovers we discuss here.

There are three main approaches to addressing the issue of complementarities and spillovers, each of which offers distinct advantages depending on the application. The first approach is using richer structural approaches that directly model complementarities and spillovers. Unfortunately, there are limited methods for dealing with asset-level complementarities in demand estimation (Berry and Haile, 2021).[3] In the context of financial markets, Allen, Kastl, and Wittwer (2025) estimate a model of demand complementarities in simultaneous auctions of treasury bonds. In contrast to equity markets, their setting

[2] See also additional studies of index inclusions (Chang, Hong, and Liskovich, 2015; Pavlova and Sikorskaya, 2023; Greenwood and Sammon, 2025), and research on fund flows (Gompers and Metrick, 2001; Coval and Stafford, 2007; Lou, 2012; Ben-David, Li, Rossi, and Song, 2022; Hartzmark and Solomon, 2024; Li, 2025) and central bank interventions (Krishnamurthy and Vissing-Jorgensen, 2011; Selgrad, 2023).

[3] In industrial organization, existing research typically considers small choice sets with limited complementarities. Gentzkow (2007) studies newspaper demand with a choice between print, online, or both.

features a small number of assets and data on bid *schedules*, not just portfolio holdings.

The second approach is to impose additional structure on the substitution matrix. For example, Koijen and Yogo (2020) and Chaudary, Fu, and Li (2023) use nested logit demand systems to study international financial markets and corporate bond markets. Nested logit allows researchers to capture specific forms of heterogeneous substitutions. In the case of corporate bonds, for instance, it is reasonable to assert that substitution within investment grade bonds is easier than across rating groups. Accordingly, Chaudary, Fu, and Li (2023) find much larger elasticities when allowing for heterogeneous substitution. A different restriction is used in An and Huber (2024), who model substitution along a small number of factors in foreign exchange, and Nenova (2025), who allows for more flexible substitution patterns using a rich set of covariates. The main limitation of these approaches is that they require ex-ante restrictions on the substitution matrix, even though substitutability is endogenously determined alongside returns and portfolios. Moreover, these restrictions must be valid for the marginal investor, and spillovers may still occur within groups of assets. In line with our findings, appropriate restrictions may be easier to ascertain for aggregate portfolios than individual securities.

The third approach is the use of reduced-form methods using control variables or difference-in-difference estimators. For example, van der Beck (2022) argues that estimating demand elasticities from trades rather than positions eliminates some endogeneity concerns in instrumental variables, while Haddad, He, Huebner, Kondor, and Loualiche (2025) show that, if one is willing to make strong symmetry assumptions on the substitution matrix across all assets in the choice set, one can identify *relative* elasticities between pairs of assets. The main limitation of this approach is that one cannot estimate the absolute elasticity, and that the required symmetry assumptions are unlikely to hold without additional conditioning information, such as control variables. As we discuss, the use of control variables generates estimates of a *residual relative* elasticity.

By allowing for tastes over assets, our paper also contributes to a growing literature in which investors hold securities because of non-pecuniary values (Starks, 2023). The literature includes the study of investment in "green assets" associated with sustainable investments.[4] Fuchs, Fukuda, and Neuhann (2025) show that "tastes" are orthogonal to returns if investors have deontological preferences, but not if they are consequentialist.

[4] See, for example, Pastor, Stambaugh, and Taylor (2021) and D'Amico, Klausmann, and Pancost (2023).

2 Framework

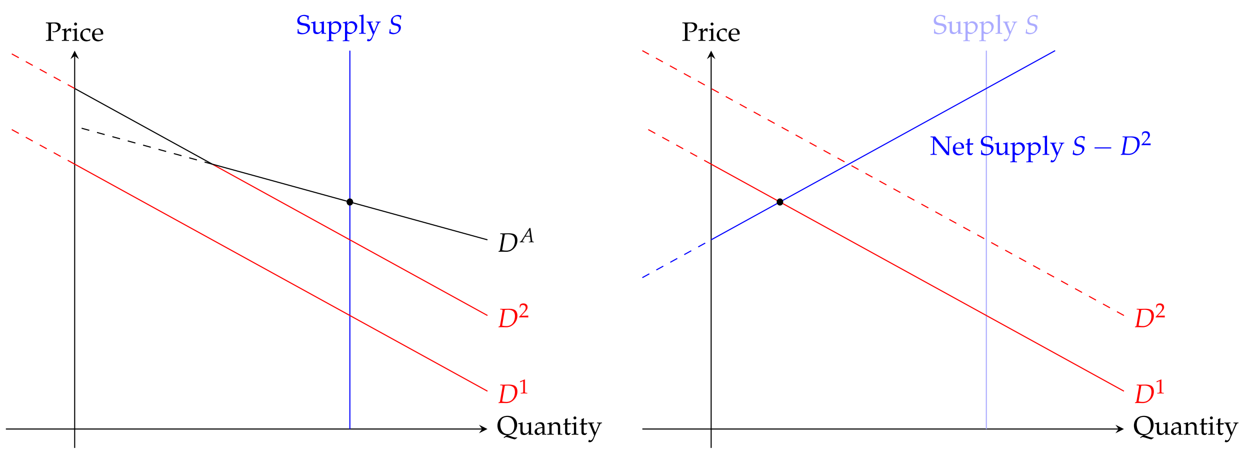
We begin by describing the general framework we use throughout the paper. Our goal is two-fold: to clarify the theoretical underpinnings of asset demand systems, and to assess whether current approaches to demand estimation can accurately identify structural elasticities. Our approach is to specify a canonical general equilibrium economy based on Lucas (1978) and to evaluate the properties of asset demand systems within this framework. To accommodate features necessary to match heterogeneous portfolios and achieve identification, we enrich this model with investor "tastes," defined as heterogeneous valuations for financial assets across different investors. We begin by reviewing the rationale for these tastes, and then formally describe our framework.

2.1 The importance of heterogeneous valuations

We start with reviewing key model ingredients necessary to allow for the identification of investor-level demand functions for financial assets from observational data. The main challenge is that quantities and prices are jointly determined in equilibrium. Hence simple regressions of quantities on prices do not identify structural parameters.

To address this issue, the literature on demand-system asset pricing focuses on variation in *net supply*, defined as aggregate supply minus demand of a subset of market participants. Figure 2 illustrates this approach. The left panel shows the canonical supply and demand diagram in an endowment economy for financial assets where the supply curve S is vertical. In the panel, D^1 and D^2 are demand curves for individual market participants, and D^A is aggregate demand. The right panel shows a potential solution: if exogenous shocks to aggregate supply S are not available, researchers can still estimate the structural parameters of demand function D^1 by finding exogenous variation in *residual* supply $S - D^2$. That is, we construct exogenous shocks to the residual supply curve of a given investor by finding exogenous shocks to the demand functions of other investors.

This approach places stringent constraints on the variation that can be used to identify demand systems. In particular, researchers must find settings in which there are demand shocks to a subset of investors that is uncorrelated with the demand of other investors. In the context of financial markets, this implies that one must find changes to market prices that are not driven by correlated shocks to discount rates and/or expected



(a) Aggregate Supply and Demand.

(b) Net Supply and Demand.

Figure 2: The basic identification issue in an endowment economy.

payoffs. Since financial assets are investment goods whose current value generically depends on their resale value, these requirements extend to expected future prices.

Two microfoundations for such shocks have been proposed. The first is cross-investor heterogeneity in *tastes* for particular assets, holding fixed a certain notion of expected cash flows. These could arise from differences in investor preferences over the provenance of cash flows, such as when some investors prefer to invest in environmentally-friendly firms. Or, investors may have dogmatic beliefs about returns that are orthogonal to the beliefs of other investors. While useful for identification, Appendix B shows that tastes can invalidate the principle of no arbitrage. The second is constraints or investment mandates that prevent some market participants from investing in a particular asset for *exogenous* reasons. We therefore use a framework that allows for both heterogeneous tastes (or dogmatic beliefs) and flexible constraints.

2.2 Formal model

We consider a one-shot portfolio choice problem in which an investor can choose to consume at date 0 and/or at date 1. Section 4.1 considers a simple dynamic extension.

There is a unit continuum of investors indexed by i. Investor i has a von Neumann-Morgenstern utility function defined over lotteries which determine the investor's consumption. A random state of the world $z \in \mathcal{Z} \equiv \{1, \ldots, Z\}$ is realized at date 1, and the probability of state z is $\pi_z \in (0, 1)$. The set of assets is $\mathcal{J} \equiv \{1, \ldots, J\}$. Asset $j \in \mathcal{J}$ offers state-contingent cash flows $y_j(z)$ in state z. Investor i is endowed with e^i_j units of asset j and additional non-asset endowment w^i_0 and $w^i_1(z)$ at dates 0 and 1, respectively. The aggregate endowment of asset j is $E_j = \int_i e^i_j di$.

Within this framework, we introduce payoff-augmenting tastes, defined as additional "consumption-equivalent" value that is generated by a particular asset. These taste parameters are designed to capture the heterogeneity in asset valuations that is required for a meaningful notion of exogenous shocks to residual demand. Formally, we say that investor i evaluates her payoffs from holding portfolio $a^i \equiv (a^i_j)_{j\in\mathcal{J}}$ by both the cash flows it generates and her *tastes* $\theta^i \equiv (\theta^i_j)_{j\in\mathcal{J}}$ over assets, where $\theta^i_j > 0$. Preferences are then defined over *effective units of consumption* delivered by portfolio a^i, and these are

$$\tilde{c}^i_1(z) \equiv \sum_{j\in\mathcal{J}} \theta^i_j y_j(z) a^i_j + w^i_1(z).$$

Investors may also be subject to portfolio constraints such as short-sale constraints, investment mandates, or restrictions on portfolio weights on particular asset classes. To capture these considerations in a flexible manner, we say that investor i faces $N \geq 0$ investment constraints on portfolio choices. The n-th investment constraint is defined as

$$M^i_n(a^i, p) \geq 0,$$

where a^i is investor i's portfolio, $p \equiv (p_j)_{j\in\mathcal{J}}$ is the price vector, and the function $M^i_n(\cdot)$ is twice continuously differentiable in a^i_j for all j. We assume that the set of feasible portfolios induced by these constraints is convex.

Given these assumptions, investor i's portfolio choice problem is:

$$\begin{aligned} \max_{a^i} \quad & (1-\delta)u^i(c^i_0) + \delta \sum_{z\in\mathcal{Z}} \pi_z u^i(\tilde{c}^i_1(z)) \\ \text{s.t.} \quad & c^i_0 = w^i_0 - \sum_{j\in\mathcal{J}} p_j(a^i_j - e^i_j) \\ & \tilde{c}^i_1(z) = \sum_{j\in\mathcal{J}} \theta^i_j y_j(z) a^i_j + w^i_1(z) \text{ for all } z \\ & M^i_n(a^i, p) \geq 0 \text{ for all } n, \end{aligned} \tag{1}$$

where $\delta \in (0,1]$ is a discount factor and the first two constraints are budget constraints at time 0 and time 1 in state z. We summarize investor i's marginal valuations by the taste-augmented marginal rate of substitution between state z and date 0, defined as

$$\tilde{\Lambda}^i(z) \equiv \frac{\delta \pi_z u^{i'}(\tilde{c}_1^i(z))}{(1-\delta) u^{i'}(c_0^i)},$$

where $u^{i'}$ is marginal utility. Our equilibrium concept is competitive equilibrium.

Definition 1 (Competitive Equilibrium) *A competitive equilibrium consists of asset prices $(p_j)_{j \in \mathcal{J}}$ and investor portfolios $(a_j^i)_{j \in \mathcal{J}}$ for each i such that:*

1. *Given asset prices, investor portfolios solve decision problem* (1) *for each i.*
2. *The consumption goods market clears in every state.*
3. *Financial markets clear for every asset: $\int_i a_j^i di = E_j$ for each $j \in \mathcal{J}$.*

2.3 Model Discussion

Our model is designed to allow for a transparent and tractable analysis of demand estimation within a canonical asset pricing framework. The main departure is the introduction of tastes parameters which generate heterogeneous asset valuations across investors. We now briefly discuss some broader implications of tastes in asset pricing.

In our approach, tastes multiplicatively augment consumption. As we will show, they operate like "latent demand" shifters in Koijen and Yogo (2019). An alternative approach is to specify additive separable tastes, whereby investor obtains some additional value (or disutility) from holding certain assets that is separable from risk-return considerations. Both formulations deliver essentially identical conclusions.

Since portfolio choice requires a cardinal interpretation of utility, the intensity of tastes influences portfolio choice. This is in contrast to many settings in industrial organization where an ordinal ranking is sufficient. Asset demand systems thus require accurate identification of the specific values of a taste parameter. Moreover, the aggregation of assets into portfolios requires appropriately weighting tastes by marginal utility.

More broadly, there is a correspondence between tastes and heterogeneous beliefs. Specifically, the state space over which payoffs are defined can be enriched to include "taste-based payoffs." Heterogeneous tastes then map into heterogeneous beliefs if investors differ in their subjective probabilities over this augmented state space. Importantly, such taste-related beliefs are dogmatic: investors agree to disagree, and in particular they may disagree on whether a particular state of the world can be realized. Such strong disagreement is *desirable* when trying to construct supply shocks because it allows for the possibility of orthogonal demand shocks. However, heterogeneous valuations are in tension with the organizing principle of no arbitrage. Appendix B shows that no arbitrage might fail given heterogeneous tastes, so that equilibrium prices may fail to exist.

3 Measuring and Interpreting Asset Demand Elasticities

In the previous section, we discussed the first main challenge of demand estimation, which is to develop conceptual frameworks for asset pricing which permit suitably exogenous shocks to residual supply. We now turn to the second main challenge, which is to develop a demand system which accurately identifies structural parameters from the data given this variation. We emphasize two critical factors that make this challenge particularly difficult in financial markets: (i) portfolio choice naturally exhibits *demand complementarities* whereby the marginal valuation of an asset depends on the rest of the investor's portfolio, and (ii) in market equilibrium, such demand complementarities generate *price spillovers* to other assets. Given these considerations, even clean exogenous variation in asset prices is generally not sufficient to identify structural parameters.

Throughout, we focus on estimating the elasticity of demand for asset j with respect to variation in some asset price p_s. This could be the asset's own price ($s = j$), or the price of another asset ($s \neq j$). Let $a^i_j(p, a^i_{-j})$ denote investor i's demand function for asset j, where p is the vector of asset prices and a^i_{-j} is the vector of investor i's remaining asset positions.[5] The elasticity of demand is the percentage change in i's demand for asset j given a percentage change in the price of asset s, *holding all other prices fixed*:

$$\mathcal{E}_{js} \equiv -\frac{\partial a^i_j(p, a^i_{-j})}{\partial p_s} \frac{p_s}{a^i_j(p, a^i_{-j})}.$$

We refer to $\mathcal{E}_{jj}$ as the own-price elasticity and to $\mathcal{E}_{js}$ ($s \neq j$) as the cross-price elasticity.

[5]Some implementations of the demand system approach define elasticities over portfolio shares rather than asset holdings, or consider demand relative to a benchmark asset. However, the appropriate units are sensitive to assumptions on investor utility and payoffs (Haddad, He, Huebner, Kondor, and Loualiche, 2025). Hence we use asset positions for now, and return to portfolio shares and relative demand later on.

3.1 The Identification Challenge

The first step is to clearly describe the identification challenge for demand systems in financial markets. To do so, we begin by deriving optimal portfolio choices (i.e., an investors' asset demand functions) from decision problem (1). Denote by λ^i_n the Lagrange multiplier associated with the n-th investment constraint, and by $m^i_{n,j}(a^i, p)$ the partial derivative of $M^i_n(a^i, p)$ with respect to a^i_j. Recall that $\tilde{\Lambda}^i(z)$ is the taste-adjusted marginal rate of substitution between date 0 and state z. Then investor i's first-order necessary condition for a^i_j, her holdings of asset j, is

$$F^i_j(a^i, p) \equiv \theta^i_j \sum_{z \in \mathcal{Z}} y_j(z)\tilde{\Lambda}^i(z) + \sum_n \lambda^i_n \frac{m^i_{n,j}(a^i, p)}{(1-\delta)u^{i\prime}(c^i_0)} - p_j = 0. \tag{2}$$

Function $F^i_j(\cdot)$ has a natural interpretation as the marginal value of asset j net of the asset price and the shadow cost of constraints. Consequently, the optimal portfolio choice is determined by the condition that the net marginal value of every asset is equal to zero,

$$F^i(a^i, p) \equiv \begin{bmatrix} F^i_1(a^i, p) \\ F^i_2(a^i, p) \\ \vdots \\ F^i_J(a^i, p) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Importantly, this system exhibits *demand complementarities* whereby the *marginal* value of asset j generically depends on the quantities held of all other assets. Hence the willingness to substitute across assets is an endogenous object that depends on the entire vector of portfolio holdings. There are two natural sources of such complementarities. The first is the canonical notion of diversification, whereby the marginal value of an asset depends on its covariance with the rest of the investor's portfolio. The second is through constraints. If an investor faces a mandate to invest at least 50% of its assets in technology stocks, buying more of any given technology stock relaxes the constraint for all non-technology stocks. In either case, optimal asset positions, and the investor's willingness to substitute between assets, are inherently intertwined with each other.

To see how these considerations complicate inference of structural parameters, assume that we have an ideal instrument in hand: a purely exogenous supply shock χ_s that directly affects only asset s. For example, in line with instrumental variable strategy of

Koijen and Yogo (2019), we might imagine that an outside investor has decided to adjust her supply of asset s for purely exogenous reasons. Using our model, we can precisely describe how the investor responds to this shock. We totally differentiate the system of the first-order conditions $F^i(a^i, p) = \mathbf{0}$ with respect to shock χ_s.

By the implicit function theorem, the optimal change of i's portfolio in response to an exogenous shock χ_s to asset s is then given by the system of equations

$$\begin{bmatrix} \frac{da_1^i}{d\chi_s} \\ \vdots \\ \frac{da_s^i}{d\chi_s} \\ \vdots \\ \frac{da_J^i}{d\chi_s} \end{bmatrix} = \begin{bmatrix} \frac{\partial a_1^i}{\partial p_1} & \cdots & \frac{\partial a_1^i}{\partial p_s} & \cdots & \frac{\partial a_1^i}{\partial p_J} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial a_s^i}{\partial p_1} & \cdots & \frac{\partial a_s^i}{\partial p_s} & \cdots & \frac{\partial a_s^i}{\partial p_J} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial a_J^i}{\partial p_1} & \cdots & \frac{\partial a_J^i}{\partial p_s} & \cdots & \frac{\partial a_J^i}{\partial p_J} \end{bmatrix} \begin{bmatrix} \frac{dp_1}{d\chi_s} \\ \vdots \\ \frac{dp_s}{d\chi_s} \\ \vdots \\ \frac{dp_J}{d\chi_s} \end{bmatrix} + \begin{bmatrix} \frac{\partial a_1^i}{\partial \chi_s} \\ \vdots \\ \frac{\partial a_s^i}{\partial \chi_s} \\ \vdots \\ \frac{\partial a_J^i}{\partial \chi_s} \end{bmatrix}. \tag{3}$$

This system formalizes the notion of cross-asset demand complementarities: while asset s is the only security that is directly affected by the shock, the observed response to the shock is the sum of endogenous quantity adjustments for *all* assets in the choice set. These adjustments reflect two channels: (i) the degree of substitutability among assets (shown in blue), which is endogenously determined from the marginal valuation of each asset given the vector of asset positions, and (ii) endogenous price spillovers to other assets (shown in red). The last term on the right-hand side captures income effects from the revaluation of endowments. Going forward, we formally denote price spillovers by

$$\mathcal{S}_{js} \equiv \frac{dp_j}{d\chi_s}.$$

The key empirical challenge is that the data does not directly reveal the structural elasticity of interest. Instead, even with exogenous price variation in hand, observational data on portfolio holdings show only the *total* derivative with respect to all margins of adjustment. We call this object the *observed elasticity* $\hat{\mathcal{E}}_{ss}^i$ and note that it is equal to

$$\hat{\mathcal{E}}_{ss}^i \equiv -\frac{\frac{da_s^i}{d\chi_s}}{\frac{dp_s}{d\chi_s}} \cdot \frac{p_s}{a_s^i}.$$

In the next result, we decompose the observed elasticity into its structural components: the sum over all assets of structural cross-elasticities multiplied by price spillovers, and the set of income effects that occur due to price changes. We focus mainly on the first component since it poses a bigger identification challenge and can lead to large biases.[6]

Proposition 1 (Decomposition of the observed elasticity) *The difference between the structural and observed own-price elasticities can be decomposed as follows:*

$$\hat{\mathcal{E}}^i_{ss} = \mathcal{E}^i_{ss} - \underbrace{\sum_{j \neq s} \frac{\mathcal{S}_{js}\frac{1}{p_j}}{\mathcal{S}_{ss}\frac{1}{p_s}} \mathcal{E}^i_{sj}}_{\substack{\text{Complementarities} \\ \text{and spillovers}}} - \underbrace{\frac{\frac{\partial a^i_s}{\partial \chi_s}\frac{1}{a^i_s}}{\mathcal{S}_{ss}\frac{1}{p_s}}}_{\text{Income effects}} . \tag{4}$$

Proof. See Appendix A. ■

To understand this result, consider again our thought experiment from the introduction, in which an investor can allocate funds between a bond and two stocks, say Microsoft and Apple. If the two stocks are highly substitutable, the cross-elasticity $\mathcal{E}^i_{MA}$ is high and investors would rapidly switch to Apple in response to a price increase for Microsoft. In equilibrium, a supply shock to Microsoft must therefore trigger a price spillover to Apple. This price spillover makes it less appealing to substitute, driving a wedge between the observed elasticity (which incorporates all adjustments) and the structural elasticity (which presumes that all other prices remain fixed). Using the observed elasticity as a measure of the structural elasticity can therefore lead to large biases.

Under heterogeneous tastes, moreover, the nature of cross-asset complementarities is investor-specific. Hence asset demand systems typically estimate substitution patterns at the level of the investor. However, there are also cross-investor interactions because price spillovers depend on the *marginal* investor's willingness to substitute across assets.

In principle, one can overcome the identification challenge by using multiple observed data moments. In particular, exogenous variation in prices allows researchers to measure observed elasticities and cross-elasticities for multiple assets. These observed elasticities can then be used to construct estimators of the true elasticity. The approach in Koijen and Yogo (2019) is to define a logit demand system in which demand is measured *relative* to an "outside asset." Under the assumption of homogeneous substitution between all assets (the standard *Independence of Irrelevant Alternatives* (IIA) property of logit demand), observed elasticities of relative demand identify the associated structural

[6] Income effects can be eliminated by redefining the units of demand. Under iso-elastic utility, for example, there are no income effects if one defines demand in terms of portfolio shares. See Section 3.2.

elasticity. More recently, Haddad, He, Huebner, Kondor, and Loualiche (2025) show that difference-in-difference estimators can be used to difference out cross-asset spillovers if the portfolio choice problem admits a symmetric substitution matrix.

The limitation of these approaches is that they are valid only under strong symmetry assumptions that are unlikely to hold in standard portfolio choice settings, at least without further conditioning information. In particular, in standard settings, substitution patterns between inside assets (i) are generically heterogeneous across different assets, both with respect to each other and the outside asset, and (ii) depend on interactions with the rest of the investors' portfolio. As such, the substitution matrix is endogenously determined alongside the asset portfolio itself. As we show next, violations of the assumption of symmetric substitution can lead to large biases in the estimated elasticities.

3.2 Biased Elasticities in Logit Demand Systems for Financial Assets

We have shown that asset demand systems can accurately recover structural elasticities only if they appropriately account for demand complementarities and price spillovers. We now show that approaches based on the logit framework from Koijen and Yogo (2019) do not satisfy this requirement for canonical portfolio choice models. As such, estimated elasticities are subject to large biases. To make this argument, we must compare estimated and structural elasticities. Our approach is to use our model to generate "data" and then ask whether current methodologies accurately identify structural model parameters.

Asset menu. Solving for demand functions in arbitrary asset menus is complicated and not necessary to make our points. Hence we consider a relatively sparse setting in which there are only three assets: two *inside assets* which are the focus of the demand estimation, and an *outside asset* towards which investors can substitute in response to shocks. This setting is rich enough to capture many cross-asset substitution patterns of practical interest, but sufficiently simple to derive closed-form expressions for many key objects of interest. Furthermore, under specific parameter restrictions, logit demand can accurately capture true model-derived demand functions.

Definition 2 (Three-asset economy) *There are two aggregate states, $z \in \{1, 2\}$, and a distributional shock $\iota \in \{r, g\}$ which further affects asset payoffs. The probability of state z is π_z, and the probability of distributional shock ι satisfies $\Pr(\iota = r) = \rho$. There are three assets:*

(i) Tree 2, which pays $y(2)$ if and only if state 2 is realized. We will refer to this asset as the outside asset, and normalize its price to $p_2 = 1$. The aggregate endowment of this asset is 1.

(ii) A green tree with price p_g which pays only in aggregate state 1, and pays more when $\iota = g$. The aggregate endowment of this asset is $\frac{1}{2} + \psi$, where ψ is an exogenous supply shifter.

(iii) A red tree with price p_r which pays only in aggregate state 1, and pays more when $\iota = r$. The aggregate endowment of this asset is $\frac{1}{2}$.

The specific state-contingent payoffs of all three assets are summarized in Table 1.

		State 1 (π_1)		State 2 ($1-\pi_1$)
		Green shock ($1-\rho$)	Red shock (ρ)	
Tree 1	green	$y(1)+\epsilon$	$y(1)-\epsilon$	0
	red	$y(1)-\epsilon$	$y(1)+\epsilon$	
Tree 2		0		$y(2)$

Table 1: Payoff structure in the three-asset economy.

We also make the following assumptions on investor endowments and constraints:

(i) Investors have the same initial endowments: $e^i_j = E_j$ and $w^i_0 = 0 = w^i_1(z)$ for all i, j and z.

(ii) Investors care only about consumption at date 1: the discount factor is $\delta = 1$.

Variation in the parameter ϵ allows us to capture a number of different scenarios. If $\epsilon = 0$, then green and red trees are perfect substitutes with respect to their cash flows. Hence these parameter values capture investors who face a security menu with similar assets, such as when they choose among similar stocks or derivative assets as well as stock. The assets become more complementary as ϵ increases. Hence intermediate values of ϵ capture when assets are complementary because they allow the investor to diversify distributional risk, but investors are still willing to substitute between assets to some degree because distributional risk is not too large. Finally, the limit $\epsilon \to y(1)$ leads to three distinct states of the world, each associated with a single tree that cannot be substituted for each other. This maps into scenarios in which there are no diversification benefits between green and red trees at all, such as when we consider an investor choosing between well-diversified portfolios each exposed to certain aggregate risk factors. Lastly, substitutability is also modulated by latent taste parameters θ^i. Formally, these are unobserved demand shifters that cannot be controlled for using data on asset cash flows alone.

Logit specification. The logit demand system in Koijen and Yogo (2019) describes asset demand in terms of portfolio shares $\omega_j^i(p) \equiv \frac{p_j a_j^i}{W^i}$, where $W^i \equiv \sum_{j \in \mathcal{J}} p_j a_j^i$ is the market value of investor i's portfolio. The *relative portfolio share* of asset j is the portfolio share of asset j divided by the portfolio share of the outside asset, $\omega_j^i(p)/\omega_2^i(p)$. Demand is specified in this manner because the logit demand system presumes that cross-asset spillovers can be appropriately controlled for measuring demand relative to the outside good.[7] The underlying logic stems from the Independence of Irrelevant Alternatives (IIA) property of logit demand, whereby substitution patterns are assumed to be homogeneous across assets. However, this assumption is at odds with the heterogenous substitution patterns that naturally arise in standard portfolio choice settings with demand complementarities.

Going forward, we adapt our definitions of elasticities to units of relative portfolio shares as well. In particular, structural and observed elasticities are now given by

$$\mathcal{E}_{jj}^i \equiv -\frac{\partial(\omega_j^i(p)/\omega_2^i(p))}{\partial p_j}\frac{p_j}{\omega_j^i(p)/\omega_2^i(p)} \qquad \text{and} \qquad \hat{\mathcal{E}}_{jj}^i \equiv -\frac{d(\omega_j^i(p)/\omega_2^i(p))}{dp_j}\frac{p_j}{\omega_j^i(p)/\omega_2^i(p)}.$$

Cross-elasticities of relative portfolio shares are defined in the analogous way.

Identification of logit demand. We first review the precise specification and identification of logit demand systems for financial assets. In this approach, relative portfolio shares are specified to be log-linear in log prices and a set of factor loadings $(x_k(j))_k$ on asset characteristics.[8] Characteristics are used to summarize key properties of expected returns and the variance-covariance matrix of returns. Specifically, in Koijen and Yogo (2019), investor-level relative portfolio shares are assumed to satisfy the demand function

$$\frac{\omega_j(p)}{\omega_2(p)} = \frac{\omega_j}{\omega_2}(p_j) = \exp\left\{\beta_0 \log p_j + \sum_{k=1}^{K-1} \beta_k x_k(j) + \beta_K\right\}\zeta(j), \tag{5}$$

where β_0 is the coefficient on the log price of asset j, $x_k(j)$ is asset j's loading on the k-th characteristics-based factor, $(\beta_k)_{k=1}^{K-1}$ are the associated demand coefficients, and β_K and

[7]Given iso-elastic utility, working with portfolio shares has the additional benefit that supply shocks do not create direct income effects. Haddad, He, Huebner, Kondor, and Loualiche (2025) provide a more general taxonomy of the "natural units" of analysis for different utility functions.

[8]Technically, β_0 is the coefficient of log market equity, where market equity is the product of price and number of shares. This ensures neutrality to variation in the level of prices that is not economically meaningful, such as different choices for the initial number of shares in a firm. However, if the number of shares is constant, this term is another additive constant. Hence we focus simply on log prices.

$\zeta(j)$ are demand shifters for all inside assets and asset j, respectively. The latent asset-specific demand parameter $\zeta(j)$ corresponds to tastes θ_j in the context of our model.

A useful geometric interpretation of (5) is that β_0 measures the slope with respect to the own price, and all other demand shifters jointly determine the intercept. The key identifying assumption is that all demand shifters other than p_j are invariant to exogenous variation in p_j. Under this assumption, one can identify slope coefficient β_0 from the observed relative elasticity. In particular, totally differentiating (5) with respect to price p_j yields $\beta_0 = \hat{\mathcal{E}}^i_{jj}$, and this parameter recovers the structural elasticity as well, $\beta_0 = \mathcal{E}^i_{jj}$. This logic is shown in the left panel of Figure 1, where the observed change in relative portfolio weights of a given asset is treated as a move *along* a fixed demand curve. Appendix C characterizes how β_0 determines the *absolute* elasticity of demand for asset j, not just the relative demand elasticity with respect to the outside good. In the context of our model, the absolute elasticity is also precisely equal to β_0.

Biased Measurement. Of course, logit demand systems fail to recover the structural elasticity if there are cross-asset spillovers and demand complementarities. This logic is depicted in the right panel of Figure 1, which incorporates the fact that a shock to one asset affects demand for other assets via complementarities and spillovers, triggering a shift in the demand curve. This effect is not captured in the logit demand system because factor loadings and characteristics, which proxy for expected returns and covariances, are assumed to be invariant to instrumented price shocks to a given asset. Given this restriction, observed elasticities are interpreted as determining the slope of demand.[9]

When complementarities exist, the observed elasticity is a biased estimator of the structural elasticity. The next result shows that the bias is directly proportional to complementarities and spillovers. Our formal definition of demand complementarities is that relative demand for asset j is affected by the prices of other assets, i.e., $\frac{\partial\left(\omega^i_j(p)/\omega^i_2(p)\right)}{\partial p_{-j}} \neq 0$ in our three-asset economy, where p_{-j} is the price of the other inside asset.

Proposition 2 (Biased measurement of structural elasticities) *Consider the three-asset econ-*

[9]This restriction is derived under the assumption that *returns* are well-described by factor loadings. However, returns are endogenous objects that should respond to price changes.

omy. For investor i, the bias $\mathcal{B}^i_{jj} \equiv \mathcal{E}^i_{jj} - \hat{\mathcal{E}}^i_{jj}$ *between structural and observed elasticities is*

$$\mathcal{B}^i_{jj} = -\underbrace{\frac{\partial\left(\omega^i_j(p)/\omega^i_2(p)\right)}{\partial p_{-j}}}_{\text{Complementarity}} \underbrace{\frac{p_{-j}}{\left(\omega^i_j(p)/\omega^i_2(p)\right)}\frac{p_j}{p_{-j}}}_{\text{Scaling terms}} \underbrace{\frac{dp_{-j}}{dp_j}}_{\text{Price Spillover}} . \tag{6}$$

Proof. See Appendix A. ■

The scaling terms ensure that the bias is reported in units of elasticities. We recover the logit identification result if and only if there are no complementarities or spillovers. Conversely, the bias from using logit demand is large when inside assets are very close substitutes (and thus much better substitutes for each other than for the outside asset). Indeed, we show formally that the bias diverges to infinity if inside assets are perfect substitutes, but zero if all assets are equally complementary.

To derive these results, we use our models to characterize structural demand elasticities under the optimal demand functions for our setting. To make these demand functions easier to interpret, we assume that there is no aggregate risk, $y(z) = 1$, although this is not necessary for any of the results to come. We have the following characterization.

Lemma 1 (Optimal portfolio shares) *Let* $y(1) = y(2) = 1$. *Given relative prices* (p_g, p_r) *and taste parameters* (θ^i_g, θ^i_r), *the optimal relative portfolio shares for investor i are given by*

$$\frac{\omega^i_g(p_g,p_r)}{\omega^i_2(p_g,p_r)} = \theta^i_r \frac{\pi_1}{\pi_2} p_g \cdot \frac{(\theta^i_r p_g + \theta^i_g p_r)\epsilon^2 - (\theta^i_r p_g - \theta^i_g p_r) + 2\theta^i_g p_r \epsilon(1-2\rho)}{(\theta^i_r p_g + \theta^i_g p_r)^2\epsilon^2 - (\theta^i_r p_g - \theta_g p_r)^2}; \tag{7}$$

$$\frac{\omega^i_r(p_g,p_r)}{\omega^i_2(p_g,p_r)} = \theta^i_g \frac{\pi_1}{\pi_2} p_r \cdot \frac{(\theta^i_r p_g + \theta^i_g p_r)\epsilon^2 + (\theta^i_r p_g - \theta^i_g p_r) - 2\theta^i_r p_g \epsilon(1-2\rho)}{(\theta^i_r p_g + \theta^i_g p_r)^2\epsilon^2 - (\theta^i_r p_g - \theta^i_g p_r)^2}. \tag{8}$$

Proof. See Appendix A. ■

The critical feature of these demand functions is that they exhibit cross-asset complementarities, whereby the demand for green and/or red assets depends on the prices and tastes of both inside assets. As such, supply shocks to one asset alter relative portfolio shares of both assets, with the degree of complementarity influenced by parameter ϵ.

The next result formally characterizes the estimation bias that obtains when using the observed elasticity as an estimator for the structural elasticity. To derive this result, we assume that we have access to a perfect asset-level instrument, namely a purely exogenous shock ψ to the supply of the green asset. While not necessary for the results, to obtain easily interpretable conditions for the bias we assume that all investors have the same tastes, $\theta^i_j = 1$, and that both aggregate states are symmetric, $\pi_1 = \frac{1}{2}$ and $y(1) = y(2) = 1$.

Proposition 3 (Measured vs Structural Elasticities) *Let $y(z) = 1 = \theta^i_j$ and $\pi_z = \frac{1}{2}$. Given an exogenous supply shock ψ around $\psi = 0$, for any i, observed and structural elasticities are:*

$$\hat{\mathcal{E}}^i_{gg} = \frac{(1-\epsilon^2)}{(1+\epsilon)^2 - 4\epsilon\rho} \qquad \textit{and} \qquad \mathcal{E}^i_{gg} = \frac{(1-\epsilon^2)(1-\epsilon(1-2\rho))}{8\rho(1-\rho)\epsilon^2}. \tag{9}$$

In the limit as green and red assets become perfect substitutes, we have:

$$\lim_{\epsilon\to 0} \hat{\mathcal{E}}^i_{gg} = 1 \qquad \textit{and} \qquad \lim_{\epsilon\to 0} \mathcal{E}^i_{gg} = \infty. \tag{10}$$

For any $\epsilon > 0$, the bias, i.e., the difference between structural and observed elasticities, is

$$\mathcal{B}^i_{gg} = \frac{(1-\epsilon^2)^2(1+(1-2\rho)\epsilon)}{8\epsilon^2\rho(1-\rho)((1+\epsilon)^2 - 4\epsilon\rho)}. \tag{11}$$

The bias is positive, goes to infinity as $\epsilon \to 0$, is strictly decreasing in ϵ, and is zero iff $\epsilon = 1$.

Proof. See Appendix A. ■

Figure 3 illustrates the result. Measured elasticities are small throughout and of the same order of magnitude as leading estimates in the demand-system literature. In contrast, the structural elasticity approaches infinity as ϵ goes to zero (such as when inside assets are closely substitutable), but is zero when $\epsilon = 1$. This means that the measured elasticity is not closely related to the underlying structural elasticity.

To understand the intuition for this disconnect, consider the case where the two inside assets are close substitutes, $\epsilon \approx 0$. (Note that, in the extreme case where $\epsilon = 0$, any change in the green asset's price leads to an immediate arbitrage opportunity relative to the red asset.) Precisely because investors would like to respond to the green supply shock by buying more of the red asset, the red price must adjust to clear the market. Given this price spillover, it becomes less attractive to substitute across assets, leading to low *measured* elasticities in response to the shock. However, the reason that prices adjust in this manner is precisely that each individual investor would have responded very elastically had prices not adjusted.

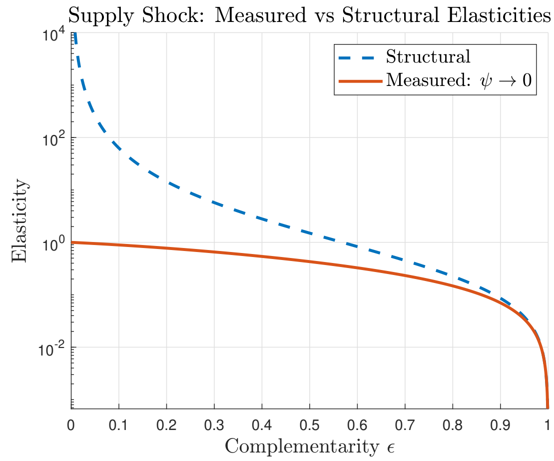


Figure 3: Measured vs structural elasticity given an exogenous supply shock ψ as $\psi \to 0$. The measured elasticity is $\hat{\mathcal{E}}^i_{gg}$. The structural elasticity is $\mathcal{E}^i_{gg}$. Parameters: $\pi_1 = \frac{1}{2}$, $y(1) = y(2) = 1$, $\theta^i_j = 1$, and $\rho = \frac{1}{4}$.

Our findings thus align directly with the basic mechanism of neoclassical finance: if financial markets rapidly reprice substitute assets in response to shocks, equilibrium portfolio responses may suggest very low elasticities even when the structural elasticity is high. The reason is that individual demand responses are strategic substitutes: when other investors adjust their portfolios, the resulting price response implies that any individual investor will rebalance her portfolio less than she otherwise would. Hence low observed elasticities are *not* dispositive evidence that financial markets are indeed slow to respond to profitable trading opportunities.

While the bias is strictly positive, it is decreasing in ϵ and converges to zero in the limit as $\epsilon \to 1$. In this limit, inside assets are not substitutable: the payoff structure features three states, each of which are associated with a single asset. Hence substitution between inside assets is symmetric to substitution with the outside asset, and demand complementarities and cross-asset spillovers are immaterial. As a result, logit demand accurately captures the underlying substitution patterns.

As such, our analysis also admits a more positive interpretation: while the bias is severe when close substitutes are available, it may be more muted when investors face choice sets with limited substitutability, as may be the case in international finance or foreign exchange (Jiang, Richmond, and Zhang, 2023; An and Huber, 2024; Koijen and Yogo, 2020).[10] Nevertheless, substitutability is ultimately a latent variable that must be esti-

[10]Limited substitutability can also stem from tastes. Investors with a taste for one asset may not reallocate to another asset in response to a price shock to their preferred asset. See Section 4.2 for an example.

mated from data. Our results suggest that logit demand cannot accurately discriminate between varying degrees of substitutability: the measured elasticity ranges from zero to one while the structural elasticity ranges from zero to infinity.

More generally, an ideal estimation procedure would control for spillovers by estimating a fixed point between individual demands and the matrix of spillovers. While theoretically appropriate, the implementation of such a fixed-point approach may be challenging. Price spillovers depend on the market-wide degree of substitutability between assets, which in turn depends on the unobserved cross-sectional distribution of taste parameters. Hence researchers would have to jointly estimate the spillover matrix alongside individual demand functions, checking for consistency using market clearing. This means that demand systems may be simpler to estimate for more aggregated portfolios, such as stocks versus bonds, where cross-asset spillovers are likely to be relatively small. However, it may also be more difficult to find instruments for such settings.

Remark 1 (Aggregation from "Micro-logit" to Asset-level Elasticities) *We have derived estimation biases at the level of the individual investor ("micro logit"). One could also derive stock-level elasticities by averaging across investors. This does not solve the identification problem: since investor-level elasticities are always underestimated, so are asset-level elasticities. In the specific setting of Proposition 3, asset-level elasticities are identical to the asset-investor-level elasticity.*

More generally, one can consider variations on our economy with heterogeneous investors. For example, some investors may have a strong taste for green assets, while the remainder have no specific taste for either asset (as in the baseline). In this case, investors with a taste-based preference for a given asset will exhibit low (or even zero) structural elasticities, which the logit demand system will identify relatively accurately. However, it will fail to accurately identify the elasticity of the investors without specific tastes. Furthermore, precisely because taste-based investors are less willing to substitute, non-taste investors will be the marginal investors whose preferences determine the relative prices of green and red assets (and thus the asset-level price response to shocks). Hence estimated asset-level demand elasticities will again be severely biased.

3.3 Control Variables

One suggested approach to the problem of heterogeneous substitution and spillovers is the use of control variables. For instance, a supply shock to a given asset may trigger spillovers primarily to other assets with similar factor exposures. In this case, controlling for common factor exposures mitigates the resulting asymmetry in substitution.

While useful for estimating certain parameters of interest, controls directly alter the object of analysis by changing the degree of substitutability between choices. In particular, two assets may be highly substitutable precisely *because* they have common exposures. In this case, controlling for common exposures yields demand elasticities defined over the *residual* cash flows unaccounted for by controls, rather than asset-level elasticities. Moreover, if residual cash flows are less substitutable than the asset, residual elasticities are lower and may carry little information about asset elasticities. Hence control-variables approaches are informative about asset demand only under additional assumptions.

We next illustrate this effect in the context of risk-based portfolio choice when asset payoffs obey a factor structure.[11] Substitutability is determined by covariances: assets are substitutable if cash flows covary positively, and complementary when they covary negatively. Hence conditional and unconditional substitutability differ whenever the conditional and unconditional covariances differ. If these differences are large, conditional demand functions are not informative about asset-level demand functions.

Example 1 (Controlling for exposures) *Let asset cash flows Y_j follow as single-factor model,*

$$Y_j = \beta_j F + \eta_j,$$

where $F \sim \mathcal{N}(\mu, \sigma^2)$ is the single factor, β_j asset j's loading on the factor and η_j is mean-zero noise that is uncorrelated across assets. Then the covariance of cash flows for two assets a and b is

$$Cov(Y_a, Y_b) = \beta_a \beta_b \sigma^2, \tag{12}$$

while the covariance conditional on the factor is zero,

$$Cov(Y_a, Y_b \mid F) = 0.$$

Hence the residual cash flows are more complementary than the asset as a whole if

$$Cov(Y_a, Y_b \mid F) < Cov(Y_a, Y_b), \quad \text{that is,} \quad \beta_a \beta_b > 0.$$

[11] Factors are common control variables because they allow researchers to hold fixed certain risk exposures. The underlying logic is based on no arbitrage pricing of risk exposures. However, allowing for heterogeneous tastes may invalidate no arbitrage: see Appendix B.

Under relatively weak assumptions, knowledge of (taste-adjusted) factor loadings may be sufficient to determine the *sign* of the difference between residual and asset-level demand elasticities. For example the residual elasticity is likely to be lower than the asset-level elasticity if $\beta_a \beta_b > 0$ and higher if $\beta_a \beta_b < 0$. However, the precise *quantitative* mapping between the two depends on a number of (unobserved) variables, including the perceived contribution to total variance of the controls, the investor's demand elasticities over the subset of asset cash flows correlated with the controls, and the interaction of each asset with the rest of the investor's portfolio. Moving from residual elasticities to asset-level elasticities therefore requires additional, potentially strong, assumptions.

In Appendix D, we formally characterize the informativeness of residual and factor-level demand elasticities in the context of our model. We decompose payoffs into factors and introduce small idiosyncratic noise. This allows us to explicitly model demand functions for factors and for residual cash flows. Under standard assumptions, factor demand elasticities are zero and residual elasticities are equal to one *independently* of the underlying asset-level elasticity. Since the asset-level elasticity ranges from zero to infinity, factor and residual demand elasticities carry little information about asset-level demand.

These findings relate to Haddad, He, Huebner, Kondor, and Loualiche (2025), who argue that conditioning information can be used to obtain settings with specific symmetry properties. Specifically, if the asset menu does not satisfy symmetry unconditionally, judicious controls may yield a decision problem that is symmetric conditional on controls. Given such symmetry, they show that a difference-in-difference estimator identifies the *relative elasticity* between two assets subject to symmetric spillovers (namely, the change in the demand difference between the two assets given a change in the price difference). However, this procedure does not recover the absolute elasticity, and it works only if substitution patterns are symmetric across *all* assets within the asset menu, as well as with outside assets. In practice, this assumption is unlikely to hold without conditioning information.[12] Using controls leads to an estimate of the *residual* relative elasticity.

[12] While their symmetry assumption is satisfied in our example economy, small perturbations of the payoff structure which break symmetry can lead to very large biases. These results are available upon request.

4 Are Elasticities Structural Parameters?

The previous section discussed the measurement and interpretation of demand elasticities, but stopped of establishing whether asset demand elasticities are structural parameters that are invariant to counterfactuals, even if they are well-measured. This is a critical concern if asset demand systems are to be used to inform policy.

We discuss two main challenges. The first is that investors care about the resale value of their assets. Hence demand elasticities reflect not only the investor's individual tastes for an asset, but also her expectations about *other* investors' future valuations. Second, it is difficult to separately identify unobserved tastes and constraints. The standard logit asset demand system cannot separately identify these different sources of demand, since all unobserved variation is summarized using a single "latent demand" parameter. However, many counterfactuals are sensitive to the precise microfoundation.

4.1 Dynamic trading: whose preferences are being measured?

We begin by discussing the role of resale considerations. Forward-looking demand makes it difficult to separately identify individual and "market-wide" tastes because investors may buy an asset they personally dislike if they expect other investors will pay a high price for it (Keynes, 1936; Harrison and Kreps, 1978). However, counterfactuals involving shifts in the wealth distribution require knowledge of individual preferences.

In Appendix E we formally construct a simple two-period variant of our baseline framework for the special case where green and red trees have identical cash flows ($\epsilon = 0$), but variation in tastes creates variation in prices. Here, we only report the key equation determining asset demand in the first period.

Let $p_{j,1}$ denote the price of asset j at date 1, R_j the gross return of asset j between dates 1 and 2, and W_2^i the investor's wealth at date 2. Given discount factor δ, an interior choice of $a_{j,1}^i$, the investor i's holdings of asset j at date 1, must satisfy

$$\frac{\pi_1 y(1)\theta_{j,1}^i}{p_{j,1}} u'\left(\tilde{c}_1^i(1)\right) = \pi_2 y(2) u'\left(\tilde{c}_1^i(2)\right) + \frac{\delta}{1-\delta}\mathbb{E}^i\left[\frac{R_2 - R_j}{W_2^i}\right]. \tag{13}$$

The left-hand side is the date-1 marginal benefit of buying asset j at current price $p_{j,1}$ in state 1. The parameter $\theta_{j,1}^i$ is investor i's taste for asset j at date 1. The right-hand side consists of two components: the marginal loss from consuming less in state 2 at date 1, and the expected return reduction from carrying wealth forward in the form of tree j rather than tree 2. Hence demand is increasing in both private tastes and expected market returns. Since market returns are determined by the tastes of tomorrow's marginal investor, observed demand elasticities do not reveal whether an investor is buying based on her tastes or her beliefs over market-wide tastes.

4.2 Tastes versus constraints

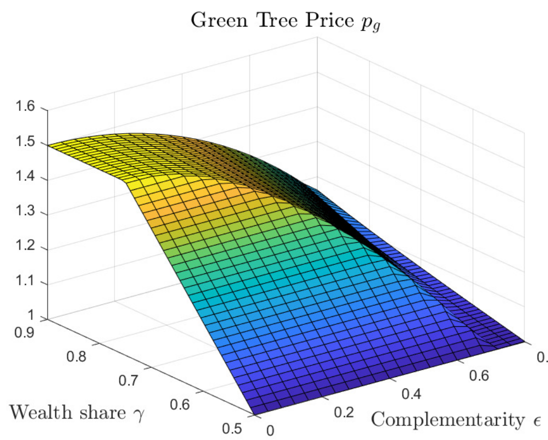
Next we consider the interpretation of observed elasticities when investors differ in both tastes and unobservable investment mandates. To this end, we study our baseline economy but assume that there are two types of investors: some prefer green to red, and the rest prefer red to green. "Green investor" owns a share γ of the aggregate endowment of every tree. This allows us to model counterfactual wealth distributions.

To ensure equilibrium existence, we assume that each investor faces short-sale constraints. Moreover, we assume that a share m of each type faces a strict mandate to only invest in their preferred trees (i.e., a green investor with a mandate cannot buy red trees). This mandate is unobserved to the econometrician. All derivations are in Appendix F.

Since investors have different tastes, the equilibrium may feature sorting. In particular, if diversification motives are sufficiently small ($\epsilon \approx 0$) and green investors are not too wealthy ($\gamma \leq \overline{\gamma}$ for some $\overline{\gamma}$), all green investors buy only green assets. In a sorting equilibrium, green investors without a mandate are observationally equivalent to green investors with a mandate. However, sorting breaks down when assets are complementary ($\epsilon \gg 0$) or when green investors are wealthy ($\gamma > \overline{\gamma}$), pushing up the green price.

Figure 4 shows the equilibrium green price p_g as a function of the wealth share γ and complementarity ϵ for two economies: one in which very few investors face a mandate (left panel), and the other with many constrained investors (right panel).

The two economies are observationally equivalent near the origin where unconstrained investors choose to specialize in their preferred color to the same extent as mandate investors. However, they differ sharply under counterfactual wealth distributions or payoff structures. For unconstrained investors, a shock to ϵ creates more demand for diversification. In the left panel, the price of green trees is thus decreasing in ϵ when green investors choose to hold inside assets. In contrast, mandate investors do not buy red trees



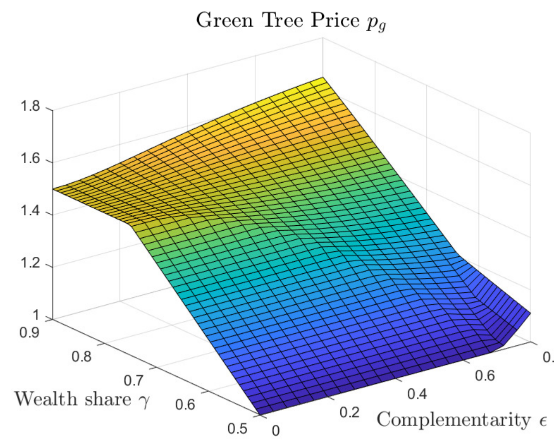


Figure 4: Green Price. Left: Mandate Share $m = 0.01$. Right: Mandate Share $m = 0.85$.

at any price. In the right panel, the price of green trees thus *increases* in ϵ. Hence, the two economies are observationally equivalent for some parameters, but qualitatively different under counterfactuals. Hence researchers should be cautious when assessing counterfactuals using demand systems that cannot separately identify tastes and constraints.

5 Conclusion

We present an analysis of demand systems for financial assets. Our results highlight the critical role of heterogeneous demand complementarities and equilibrium price spillovers. Specifically, we show that a failure to properly account for these effects can lead to severe biases in measured elasticities. This offers a simple reconciliation of the striking difference between demand elasticities in demand-system approaches and canonical benchmarks.

We see two main paths for future work. The first is to develop new structural frameworks that can account for richer cross-asset interactions. This could involve finding applications with additional data or structure, as in Allen, Kastl, and Wittwer (2025). The second is to develop empirical tests to bound the estimation biases that we discuss.

References

ADRIAN, T., E. ETULA, AND T. MUIR (2014): "Financial Intermediaries and the Cross-Section of Asset Returns," *Journal of Finance*, 69(6), 2558–2596. [Cited on page 4.]

ALLEN, J., J. KASTL, AND M. WITTWER (2025): "Estimating Demand Systems with Bidding Data," *Working Paper*. [Cited on pages 5 and 27.]

AN, Y., AND A. W. HUBER (2024): "Demand Propagation Through Traded Risk Factors," Discussion paper. [Cited on pages 6 and 21.]

BEN-DAVID, I., J. LI, A. ROSSI, AND Y. SONG (2022): "What Do Mutual Fund Investors Really Care About?," *Review of Financial Studies*, 35(4), 1723–1774. [Cited on page 5.]

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

BRECKENFELDER, J., AND V. DE FALCO (2023): "Investor Heterogeneity and Large-Scale Asset Purchases," *Working paper, ECB and Imperial College*. [Cited on page 5.]

BRETSCHER, L., L. SCHMID, I. SEN, AND V. SHARMA (2022): "Institutional Corporate Bond Pricing," *Working Paper*. [Cited on page 5.]

CHANG, Y.-C., H. HONG, AND I. LISKOVICH (2015): "Regression Discontinuity and the Price Effects of Stock Market Indexing," *Review of Financial Studies*, 28(1), 212–246. [Cited on page 5.]

CHAUDARY, M., Z. FU, AND J. LI (2023): "Corporate Bond Multipliers: Substitutes Matter," *Columbia University Working Paper*. [Cited on page 6.]

COVAL, J., AND E. STAFFORD (2007): "Asset Fire Sales (and Purchases) in Equity Markets," *Journal of Financial Economics*, 86(2), 479–512. [Cited on page 5.]

D'AMICO, S., J. KLAUSMANN, AND A. A. PANCOST (2023): "The Benchmark Greenium," *Working Paper*. [Cited on page 6.]

DARMOUNI, O., K. Y. SIANI, AND K. XIAO (2023): "Nonbank Fragility in Credit Markets: Evidence from a Two-Layer Asset Demand System," *Columbia University and MIT Working paper*. [Cited on page 5.]

DAVIS, C., M. KARGAR, AND J. LI (2025): "Why Do Portfolio Choice Models Predict Inelastic Demand?," *Available at SSRN: https://ssrn.com/abstract=4195089 or http://dx.doi.org/10.2139/ssrn.4195089*. [Cited on page 5.]

FUCHS, W., S. FUKUDA, AND D. NEUHANN (2025): "Kantian versus Consequentialist Portfolio Choice," *Working paper, University of Texas at Austin and Bocconi University*. [Cited on page 6.]

GABAIX, X., AND R. KOIJEN (2020): "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," *mimeo Harvard and Chicago*. [Cited on page 5.]

GENTZKOW, M. (2007): "Valuing New Goods in a Model with Complementarity: Online Newspapers," *American Economic Review*, 97(3), 713–744. [Cited on page 5.]

GOMPERS, P. A., AND A. METRICK (2001): "Institutional Investors and Equity Prices," *Quarterly Journal of Economics*, 116(1), 229–259. [Cited on page 5.]

GREENWOOD, R. M., AND M. SAMMON (2025): "The Disappearing Index Effect," *Journal of Finance*, 80(2), 657–698. [Cited on page 5.]

HADDAD, V., Z. HE, P. HUEBNER, P. KONDOR, AND E. LOUALICHE (2025): "Causal Inference for Asset Pricing," Discussion paper. [Cited on pages 6, 11, 15, 17, and 24.]

HADDAD, V., P. HUEBNER, AND E. LOUALICHE (2025): "How Competitive is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing," *American Economic Review*, 115(3), 975–1018. [Cited on page 5.]

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 4.]

HARRISON, J. M., AND D. M. KREPS (1978): "Speculative Investor Behavior in a Stock Market with Heterogenous Expectations," *Quarterly Journal of Economics*, 92(2), 323–336. [Cited on page 25.]

HARTZMARK, S. M., AND D. H. SOLOMON (2024): "Marketwide Predictable Price Pressure," *Working Paper, Boston College*. [Cited on page 5.]

HE, Z., AND A. KRISHNAMURTHY (2013): "Intermediary Asset Pricing," *American Economic Review*, 103(2), 732–70. [Cited on page 4.]

JIANG, Z., R. RICHMOND, AND T. ZHANG (2023): "A Portfolio Approach to Global Imbalances," *NYU Stern School of Business Working Paper*. [Cited on pages 5 and 21.]

KEYNES, J. M. (1936): *The General Theory of Employment, Interest, and Money*. Macmillan, London. [Cited on pages 4 and 25.]

KOIJEN, R. S., AND M. YOGO (2020): "Exchange Rates and Asset Prices in a Global Demand System," *NBER Working Paper 27342*. [Cited on pages 6 and 21.]

KOIJEN, R. S. J., R. J. RICHMOND, AND M. YOGO (2024): "Which Investors Matter for Equity Valuations and Expected Returns?," *Review of Economic Stdies*, 91(4), 2387–2424. [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 1, 2, 3, 10, 13, 14, 15, 17, 36, and 41.]

——— (2021): "Asset Demand Systems in Macro-Finance," *NBER Reporter*, 4, 21–24. [Cited on page 1.]

KRISHNAMURTHY, A., AND A. VISSING-JORGENSEN (2011): "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy," *Brookings Papers on Economic Activity*, 42(2), 215–287. [Cited on page 5.]

——— (2012): "The Aggregate Demand for Treasury Debt," *Journal of Political Economy*, 120(2), 233–267. [Cited on page 4.]

LI, Z. (2025): "Long Rates, Life Insurers, and Credit Spreads," *Working paper, Imperial College*. [Cited on page 5.]

LOU, D. (2012): "A Flow-Based Explanation for Return Predictability," *Review of Financial Studies*, 25(12), 3457–3489. [Cited on page 5.]

LUCAS, R. E. J. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46(6), 1429–1445. [Cited on pages 1 and 7.]

NENOVA, T. (2025): "Global or Regional Safe Assets: Evidence from Bond Substitution Patterns," *London Business School Working Paper*. [Cited on pages 5 and 6.]

PASTOR, L., R. F. STAMBAUGH, AND L. A. TAYLOR (2021): "Sustainable investing in equilibrium," *Journal of Financial Economics*, 142(2), 550–571. [Cited on page 6.]

PAVLOVA, A., AND T. SIKORSKAYA (2023): "Benchmarking Intensity," *Review of Financial Studies*, 36(3), 859–903. [Cited on page 5.]

ROSS, S. A. (2004): *Neoclassical Finance*. Princeton University Press, Princeton, NJ. [Cited on pages 4 and 35.]

SELGRAD, J. (2023): "Testing the Portfolio Rebalancing Channel of Quantitative Easing," Working paper. [Cited on page 5.]

SHLEIFER, A. (1986): "Do Demand Curves for Stocks Slope Down?," *Journal of Finance*, 41(3), 579–590. [Cited on page 4.]

STARKS, L. T. (2023): "Presidential Address: Sustainable Finance and ESG Issues - Value versus Values," *Journal of Finance*, 78(4), 1837–1872. [Cited on page 6.]

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

VAN DER BECK, P. (2022): "On the Estimation of Demand-Based Asset Pricing Models," *Swiss Finance Institute Research Paper Series*, 22(67). [Cited on page 6.]

A Proofs

Proof of Proposition 1. First, it follows from (the s-th row of) equation (3) that

$$\frac{da_s^i}{d\chi_s} = \frac{\partial a_s^i}{\partial p_s}\frac{dp_s}{d\chi_s} + \sum_{j\neq s}\frac{\partial a_s^i}{\partial p_j}\frac{dp_j}{d\chi_s} + \frac{\partial a_s^i}{\partial \chi_s}.$$

Second, multiplying both sides by $-\frac{1}{\frac{dp_s}{d\chi_s}}\frac{p_s}{a_s^i}$ and noting the definitions of $\hat{\mathcal{E}}_{ss}^i$, $\mathcal{E}_{ss}^i$, $\mathcal{S}_{js}$, and $\mathcal{S}_{ss}$, we have:

$$\hat{\mathcal{E}}_{ss}^i = \mathcal{E}_{ss}^i - \left(\sum_{j\neq s}\frac{\partial a_s^i}{\partial p_j}\frac{\mathcal{S}_{js}}{\mathcal{S}_{ss}}\frac{p_s}{a_s^i} + \frac{\frac{\partial a_s^i}{\partial \chi_s}}{\mathcal{S}_{ss}}\frac{p_s}{a_s^i}\right).$$

Third, then, rearranging, we obtain equation (4). ■

Proof of Proposition 2. Observe that the structural elasticity can be rewritten as:

$$\mathcal{E}_{jj}^i = -\frac{\partial \log\left(\frac{\omega_j^i(p)}{\omega_2^i(p)}\right)}{\partial p_j}p_j = -\left(\frac{\partial \omega_j^i(p)}{\partial p_j}\frac{p_j}{\omega_j^i} - \frac{\partial \omega_2^i(p)}{\partial p_j}\frac{p_j}{\omega_2^i}\right).$$

In contrast, the measured elasticity is written as:

$$\begin{aligned}\hat{\mathcal{E}}_{jj}^i &= -\frac{d \log\left(\frac{\omega_j^i(p)}{\omega_2^i(p)}\right)}{dp_j}p_j = -\left(\frac{d\omega_j^i(p)}{dp_j}\frac{p_j}{\omega_j^i} - \frac{d\omega_2^i(p)}{dp_j}\frac{p_j}{\omega_2^i}\right)\\ &= -\left(\frac{\partial \omega_j^i(p)}{\partial p_j}\frac{p_j}{\omega_j^i} - \frac{\partial \omega_2^i(p)}{\partial p_{-j}}\frac{p_j}{\omega_2^i}\right) - \left(\frac{\partial \omega_j^i(p)}{\partial p_{-j}}\frac{p_j}{\omega_j^i} - \frac{\partial \omega_2^i(p)}{\partial p_{-j}}\frac{p_j}{\omega_2^i}\right)\frac{dp_{-j}}{dp_j}.\end{aligned}$$

Hence,

$$\begin{aligned}\mathcal{B}_{jj}^i = \mathcal{E}_{jj}^i - \hat{\mathcal{E}}_{jj}^i &= -\left(\frac{\partial \omega_j^i(p)}{\partial p_{-j}}\frac{p_{-j}}{\omega_j^i} - \frac{\partial \omega_2^i(p)}{\partial p_{-j}}\frac{p_{-j}}{\omega_2^i}\right)\frac{p_j}{p_{-j}}\frac{dp_{-j}}{dp_j}\\ &= -\frac{\partial\left(\omega_j^i(p)/\omega_2^i(p)\right)}{\partial p_{-j}}\frac{p_{-j}}{\left(\omega_j^i(p)/\omega_2^i(p)\right)}\frac{p_j}{p_{-j}}\frac{dp_{-j}}{dp_j},\end{aligned}$$

as desired. ■

Proof of Lemma 1. Observe that, under $y(1) = y(2) = 1$, the first-order conditions with respect to a_g^i and a_r^i can be rewritten as

$$p_g = \frac{\pi_1(1-\rho)}{\pi_2}\frac{(1+\epsilon)\theta_g^i a_2^i}{(1+\epsilon)\theta_g^i a_g^i + (1-\epsilon)\theta_r^i a_r^i} + \frac{\pi_1\rho}{\pi_2}\frac{(1-\epsilon)\theta_g^i a_2^i}{(1-\epsilon)\theta_g^i a_g^i + (1+\epsilon)\theta_r^i a_r^i}; \tag{14}$$

$$p_r = \frac{\pi_1(1-\rho)}{\pi_2}\frac{(1-\epsilon)\theta_r^i a_2^i}{(1+\epsilon)\theta_g^i a_g^i + (1-\epsilon)\theta_r^i a_r^i} + \frac{\pi_1\rho}{\pi_2}\frac{(1+\epsilon)\theta_r^i a_2^i}{(1-\epsilon)\theta_g^i a_g^i + (1+\epsilon)\theta_r^i a_r^i}. \tag{15}$$

Substituting the budget constraint

$$a_2^i = p_g e_g^i + p_r e_r^i + e_2^i - p_g a_g^i - p_r a_r^i$$

and solving for (a_g^i, a_r^i), we obtain:

$$a_g^i = \theta_r^i \pi_1 \cdot \frac{(p_g e_g^i + p_r e_r^i + e_2^i)\Big((\theta_r^i p_g + \theta_g^i p_r)\epsilon^2 - (\theta_r^i p_g - \theta_g^i p_r) + 2\theta_g^i p_r \epsilon(1-2\rho)\Big)}{(\theta_r^i p_g + \theta_g^i p_r)^2\epsilon^2 - (\theta_r^i p_g - \theta_g^i p_r)^2}; \tag{16}$$

$$a_r^i = \theta_g^i \pi_1 \cdot \frac{(p_g e_g^i + p_r e_r^i + e_2^i)\Big((\theta_r^i p_g + \theta_g^i p_r)\epsilon^2 + (\theta_r^i p_g - \theta_g^i p_r) - 2\theta_r^i p_g \epsilon(1-2\rho)\Big)}{(\theta_r^i p_g + \theta_g^i p_r)^2\epsilon^2 - (\theta_r^i p_g - \theta_g^i p_r)^2}. \tag{17}$$

We then obtain the portfolio weights (ω_g^i, ω_r^i) in equations (7) and (8), respectively. ■

Proof of Proposition 3. First, observe that the representative agent (thus we suppress the superscript i) consumes the aggregate endowment in equilibrium. Thus, under the assumptions that $y(1) = y(2) = 1$, $\pi_1 = \frac{1}{2}$, and no tastes, the first-order conditions yield:

$$p_g = (1-\rho)\frac{1+\epsilon}{1+(1+\epsilon)\psi} + \rho\frac{1-\epsilon}{1+(1-\epsilon)\psi};$$
$$p_r = (1-\rho)\frac{1-\epsilon}{1+(1+\epsilon)\psi} + \rho\frac{1+\epsilon}{1+(1-\epsilon)\psi}.$$

Then, the change in prices $\Delta p_g \equiv p_g - p_g^0$ satisfies:

$$\Delta p_g \equiv p_g - p_g^0 = -\frac{1+\epsilon^2 + (1-\epsilon^2)\psi + (1-2\rho)(2+(1-\epsilon^2)\psi)\epsilon}{(1+(1+\epsilon)\psi)(1+(1-\epsilon)\psi)}\psi.$$

When $\epsilon = 0$, we have $p_g = \frac{1}{1+\psi}$, $p_g^0 = 1$, and $\Delta p_g = -\frac{\psi}{1+\psi}$.

Second, substituting the equilibrium prices into the portfolio weight function ω_g

yields the equilibrium portfolio weights of green tree:

$$\omega_g = \frac{1+2\psi}{4} \cdot \frac{1+\psi+\epsilon(1-2\rho-\epsilon\psi)}{(1+(1+\epsilon)\psi)(1+(1-\epsilon)\psi)};$$
$$\omega_g^0 = \frac{1}{4}\left(1+(1-2\rho)\epsilon\right);$$
$$\Delta\omega_g \equiv \omega_g - \omega_g^0 = \frac{1}{4} \cdot \frac{(1-\epsilon^2)\psi(1+\psi-\epsilon(1-2\rho)\psi)}{(1+(1-\epsilon)\psi)(1+(1+\epsilon)\psi)}.$$

In the limit as $\epsilon = 0$, we have $\omega_g = \frac{1+2\psi}{4(1+\psi)}$, $\omega_g^0 = \frac{1}{4}$, and $\Delta\omega_g = \frac{\psi}{4(1+\psi)}$.

Third, the measured elasticity reduces to

$$-\frac{\Delta\omega_g}{\Delta p_g}\frac{p_g^0}{\omega_g^0} = \frac{(1-\epsilon^2)(1+\psi-\epsilon(1-2\rho)\psi)}{1+\epsilon^2+(1-\epsilon^2)\psi+(1-2\rho)(2+(1-\epsilon^2)\psi)\epsilon}. \tag{18}$$

In the limit as $\psi \to 0$, we obtain the first equation in (9). Also, in the limit as $\epsilon \to 0$, the measured elasticity tends to the first equation in (10).

Forth, to compute the structural elasticity, we can compute

$$-\frac{\partial\omega_g}{\partial p_g}\frac{p_g}{\omega_g}$$

as a function of (p_g, p_r) when $(e_g, e_r, e_2) = (\frac{1}{2}+\psi, \frac{1}{2}, 1)$. Note that, since ω_g does not depend on (e_g, e_r, e_2), this term does not depend on (e_g, e_r, e_2), i.e., ψ. Then, evaluating this at the initial equilibrium prices (p_g^0, p_r^0), the structural elasticity is given by the second equation in (9), which goes to infinity as $\epsilon \to 0$, establishing the second equation in (10).

Fifth, as in Proposition 1, we can decompose the measured demand response (in the limit as $\psi \to 0$) as the true response and the bias term:

$$\frac{\frac{d\omega_g}{d\psi}}{\frac{d\log p_g}{d\psi}} = \frac{\frac{\partial\omega_g}{\partial\log p_g}\frac{d\log p_g}{d\psi} + \frac{\partial\omega_g}{\partial\log p_r}\frac{d\log p_r}{d\psi}}{\frac{d\log p_g}{d\psi}} = \frac{\partial\omega_g}{\partial\log p_g} + \frac{\partial\omega_g}{\partial p_r}\frac{\frac{dp_r}{d\psi}}{\frac{d\log p_g}{d\psi}}.$$

Then, we can decompose the measured elasticity (in the limit as $\psi \to 0$) into the structural elasticity and the bias term:

$$-\frac{\frac{d\omega_g}{d\psi}}{\frac{d\log p_g}{d\psi}}\frac{p_g}{\omega_g} = -\frac{\partial\omega_g}{\partial\log p_g}\frac{p_g}{\omega_g} - \frac{\partial\omega_g}{\partial p_r}\frac{\frac{dp_r}{d\psi}}{\frac{d\log p_g}{d\psi}}\frac{p_g}{\omega_g}.$$

Consequently, the measured elasticity in the limit as $\psi \to 0$ is:

$$-\frac{\frac{d\omega_g}{de_g}}{\frac{dp_g}{de_g}}\frac{p_g}{w_g} = \frac{(1-\epsilon^2)}{(1+\epsilon)^2 - 4\epsilon\rho},$$

which coincides with the limit of equation (18) as $\psi \to 0$. In the limit $\psi \to 0$, the bias is:

$$\frac{\partial \omega_g}{\partial p_r}\frac{\frac{dp_r}{d\psi}}{\frac{d\log p_g}{d\psi}}\frac{p_g}{\omega_g} = \frac{(1-\epsilon^2)^2(1+(1-2\rho)\epsilon)}{8\epsilon^2\rho(1-\rho)((1+\epsilon)^2-4\epsilon\rho)}. \tag{19}$$

It follows from equation (19) that the bias is positive (when $\epsilon < 1$) and goes to infinity as $\epsilon \to 0$. Also, the derivative of the bias term with respect to ϵ is

$$-\frac{(1+\epsilon)^4 + 2\epsilon(1-\epsilon)(1+\epsilon)^2(7+\epsilon)(1+\epsilon^2)\rho + 16\epsilon^2(1-\epsilon^4)\rho^2}{8\epsilon^3\rho(1-\rho)\left((1+\epsilon)^2-4\epsilon\rho\right)^2} < 0,$$

which establishes that the bias is strictly decreasing in ϵ. ∎

B No Arbitrage with Tastes

In many applications, researchers aim to control for certain risk exposures or use a low-dimensional factor representation of the matrix of expected returns to model demand. No arbitrage is used to ensure that assets are priced according to their risk exposures. We now discuss the relation between taste heterogeneity and the principle of no arbitrage.

In the standard definition, investors care only about cash flows and an arbitrage is "an investment strategy that guarantees a positive payoff in some contingency with no possibility of a negative payoff and no initial net investment" (Ross, 2004). When investors differ in tastes, they have subjective views on the payoffs of a given trade.

To define no arbitrage with tastes, we provide two preliminary definitions. First, letting a vector subspace $\mathcal{A}$ of $\mathbf{R}^J$ denote the set of feasible portfolios and letting $(p_j)_{j\in\mathcal{J}}$ be a price vector of individual assets, the pricing function $P : \mathcal{A} \to \mathbf{R}$ maps a portfolio $a = (a_j)_{j\in\mathcal{J}}$ into its price according to $P(a) \equiv \sum_{j\in\mathcal{J}} p_j a_j$. Second, investor i has a linear taste function $v^i : \mathcal{A} \to \mathbf{R}^Z$ that maps a portfolio a into a vector $v^i(a)$ of state-contingent taste-augmented payoffs for investor i. Specifically, letting $Y^i \equiv (\theta^i_j y_j(z))_{z,j}$ be the $Z \times J$ matrix of investor i's payoff-augmented cash flows, $v^i(a) \equiv Y^i a$. This generalizes the neoclassical approach in which $\theta^i_j = 1$ for all j. We then have the following definition.

Definition 3 (No Arbitrage with Tastes) *Let taste functions v^i be given for all investors i. The pricing function P leaves no arbitrage opportunities if, for any investor i and any portfolio $a \in \mathcal{A}$ such that the effective payoff is weakly positive almost surely (i.e., $v^i(a) \geq 0$) and strictly positive with strictly positive probability (i.e., $v^i(a) > 0$), the associated price is positive: $P(a) > 0$.*

That is, pricing function P leaves no arbitrage opportunities given taste functions $(v^i)_i$ if and only if, for every i, the pricing function P leaves no arbitrage opportunities in the standard sense if the cash-flow matrix is replaced by the taste-augmented payoff matrix. The key difficulty is that this payoff matrix is investor-specific. The main restriction is that taste functions are linear, as they are in Koijen and Yogo (2019).

Theorem 1 (Generic Arbitrage Opportunities with Tastes) *Fix taste functions v^i for all i. There does not exist a pricing function P that leaves no arbitrage opportunities if:*

$$\textit{there exist } a, i, \textit{ and } i' \textit{ such that } v^i(a) > 0 \textit{ and } 0 \geq v^{i'}(a). \tag{C}$$

A sufficient condition for (C) is that there exist assets j and j' such that

(i) both assets have identical cash flows:

$$y_j(z) = y_{j'}(z) \textit{ for all } z \in \mathcal{Z};$$

(ii) there exist investors i and i' with sufficiently heterogeneous tastes with respect to these assets:

$$\theta^i_j > \theta^i_{j'} \textit{ and } \theta^{i'}_j \leq \theta^{i'}_{j'}.$$

Proof. We first show that there does not exist a pricing function P that leaves no arbitrage opportunities if (C) holds. We then establish the stated sufficiency condition for (C).

Suppose first that (C) holds. Generically, we can assume that $v^i(a) > 0$ and $v^{i'}(a) < 0$. Now, suppose to the contrary that there exists a pricing function P that leaves no arbitrage opportunities. Applying Definition 3 to investor i yields $P(a) > 0$. But applying Definition 3 to investor i' yields $P(-a) > 0$, i.e., $P(a) < 0$. This is a contradiction.

Next, suppose that the stated conditions hold: there exist two assets j and j' and two investors i and i' satisfying the two conditions. Denoting by $v^i_j \equiv (\theta^i_j y_j(z))_z$ investor

i's marginal taste with respect to asset j, the two conditions imply:

$$v^i_j > v^i_{j'} \text{ and } v^{i'}_j \leq v^{i'}_{j'}. \tag{C'}$$

It is then sufficient to show that (C′) implies condition (C). Let $a \in \mathbf{R}^J$ be such that

$$a_\ell = \begin{cases} 1 & \text{if } \ell = j \\ -1 & \text{if } \ell = j' \\ 0 & \text{otherwise} \end{cases}.$$

Then, we obtain $v^i(a) = v^i_j - v^i_{j'} > 0$ and $v^{i'}(a) = v^{i'}_j - v^{i'}_{j'} \leq 0$, as desired. ■

The following example provides a simple illustration consistent with our model.

Example 2 (Green and Red Assets) *There are a green asset and a red asset with prices denoted by p_g and p_r, respectively. Both assets deliver a unit payoff with certainty. There are two investor types, denoted by α and β, that differ in their relative taste for the two assets. For each investor type i, the taste function is given by $v^i(a_g, a_r) = \theta^i_g a_g + \theta^i_r a_r$ with the following properties: while type α's taste-augmented payoffs for green and red assets satisfy $\theta^\alpha_g > \theta^\alpha_r$, type β has $\theta^\beta_g < \theta^\beta_r$. Then there are no prices such that both investors agree on the value of a long-short portfolio selling one unit of the green asset and buying one unit of the red asset.*

C Computing the absolute elasticity given β_0

In the baseline logit model, the key estimation equation is written in terms of portfolio shares relative to the outside good. We now show how β_0 can be used to back out the absolute elasticity of the portfolio share of asset j, $\partial \log(\omega_j)/\partial \log(p_j)$ given knowledge of the demand for the outside asset.

Consider $J-1$ inside assets and one outside asset. Let o denote the outside asset. For any inside assets j and k with $j \neq k$, the logit demand system assumes that

$$0 = \frac{\partial}{\partial \log p_k} \log\left(\frac{\omega_j}{\omega_o}\right)(p) = \frac{\partial \log \omega_j(p)}{\partial \log p_k} - \frac{\partial \log \omega_o(p)}{\partial \log p_k}.$$

Thus, we obtain:

$$\frac{\partial \log \omega_j(p)}{\partial \log p_k} = \frac{\partial \log \omega_o(p)}{\partial \log p_k}.$$

This equation means that, for the j-th row of the elasticity matrix $\left(\frac{\partial \log \omega_j(p)}{\partial \log p_k}\right)_{j,k}$, the (j,k)-th element (with $k \neq j$) is

$$\frac{\partial \log \omega_o(p)}{\partial \log p_k},$$

which does not depend on j. For the (j,j)-element of the elasticity matrix, we have

$$\frac{\partial}{\partial \log p_j} \log\left(\frac{\omega_j}{\omega_o}\right)(p) = \frac{\partial \log \omega_j(p)}{\partial \log p_j} - \frac{\partial \log \omega_o(p)}{\partial \log p_j}.$$

Under the assumption of the logit demand system that

$$\frac{\partial}{\partial \log p_j} \log\left(\frac{\omega_j}{\omega_o}\right)(p) = \beta_0,$$

we have:

$$\frac{\partial \log \omega_j(p)}{\partial \log p_j} = \beta_0 + \frac{\partial \log \omega_o(p)}{\partial \log p_j}.$$

We then obtain the following elasticity matrix:

$$\begin{bmatrix} \beta_0 + \frac{\partial \log \omega_o(p)}{\partial \log p_1} & \frac{\partial \log \omega_o(p)}{\partial \log p_j} & \cdots\cdots & \frac{\partial \log \omega_o(p)}{\partial \log p_{J-1}} \\ \frac{\partial \log \omega_o(p)}{\partial \log p_1} & \beta_0 + \frac{\partial \log \omega_o(p)}{\partial \log p_2} & \cdots\cdots & \frac{\partial \log \omega_o(p)}{\partial \log p_{J-1}} \\ \frac{\partial \log \omega_o(p)}{\partial \log p_1} & \frac{\partial \log \omega_o(p)}{\partial \log p_2} & \cdots\cdots & \frac{\partial \log \omega_o(p)}{\partial \log p_{J-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \log \omega_o(p)}{\partial \log p_1} & \frac{\partial \log \omega_o(p)}{\partial \log p_2} & \cdots\cdots & \beta_0 + \frac{\partial \log \omega_o(p)}{\partial \log p_{J-1}} \end{bmatrix}.$$

Letting ϖ_j be such that $\frac{\partial \log \omega_o(p)}{\partial \log p_j} = -\varpi_j \beta_0$, the elasticity matrix reduces to:

$$\begin{bmatrix} (1-\varpi_1)\beta_0 & -\varpi_2\beta_0 & \cdots\cdots & -\varpi_{J-1}\beta_0 \\ -\varpi_1\beta_0 & (1-\varpi_2)\beta_0 & \cdots\cdots & -\varpi_{J-1}\beta_0 \\ -\varpi_1\beta_0 & -\varpi_2\beta_0 & \cdots\cdots & -\varpi_{J-1}\beta_0 \\ \vdots & \vdots & \ddots & \vdots \\ -\varpi_1\beta_0 & -\varpi_2\beta_0 & \cdots\cdots & (1-\varpi_{J-1})\beta_0 \end{bmatrix}.$$

Given information on demand elasticities for the outside asset, we can then compute the elasticity in absolute terms. In the three-asset economy, the elasticity matrix is:

$$\begin{bmatrix} \beta_0 + \frac{\partial \log \omega_2(p)}{\partial \log p_g} & \frac{\partial \log \omega_2(p)}{\partial \log p_r} \\ \frac{\partial \log \omega_2(p)}{\partial \log p_g} & \beta_0 + \frac{\partial \log \omega_2(p)}{\partial \log p_r} \end{bmatrix} = \begin{bmatrix} \left(1 - \frac{-\frac{\partial \log \omega_2(p)}{\partial \log p_g}}{\beta_0}\right)\beta_0 & -\frac{-\frac{\partial \log \omega_2(p)}{\partial \log p_r}}{\beta_0} \\ -\frac{-\frac{\partial \log \omega_2(p)}{\partial \log p_g}}{\beta_0} & \left(1 - \frac{-\frac{\partial \log \omega_2(p)}{\partial \log p_r}}{\beta_0}\right)\beta_0 \end{bmatrix}.$$

In our model, we can also directly compute the optimal demand for the outside asset. Under log utility, it is trivial that $\omega_2 = \pi_2$. Hence the the elasticity matrix is

$$\begin{bmatrix} \beta_0 & 0 \\ 0 & \beta_0 \end{bmatrix}.$$

D Residual and Factor Demand Elasticities

In this appendix, we formally characterize the difference between asset-level elasticities and *residual* elasticities in the context of our model. We will show that neither factor-level demand functions nor residual demand functions accurately inform asset-level demand elasticities. To ensure that no arbitrage holds (as is required for control variables based on factor exposures), we assume that investors do not have heterogeneous tastes, $\theta_j^i = 1$.

We begin by decomposing the payoff structure from Table 1 into underlying factors. For pedagogical purposes, assume that the factors are Arrow securities on the states. Since there are two aggregate states and a distributional shock, the set of states is $\{g, r, 2\}$. By a slight abuse of notation, use z to index a generic state. Given the payoff structure from Table 1, asset j has factor loading $y_j(z)$ on the state-z Arrow security, and factor loadings perfectly capture asset payoffs. Since green and red trees can be combined to replicate Arrow securities, all factors are traded. By no arbitrage, the factors thus have well-defined prices $q(z)$ and they are related to asset prices by

$$p_j = \sum_z y_j(z) q(z).$$

Factor demands. We first show that *factor-level* demand elasticities are not useful for capturing asset-level elasticities. Since the factors are traded, we can define a decision problem directly over state-contingent consumption $c(z)$ with associated price $q(z)$. There are two notions of factor demand functions. In the first, we simultaneously choose demand functions for all factors. This corresponds to the following decision problem:

$$\begin{aligned} \max_{c(r),c(g)} \quad & \pi_2 u(c(2)) + \pi_1 \rho u(c(r)) + \pi_1(1-\rho)u(c(g)) \qquad \text{(Factor demand)} \\ \text{s.t.} \quad & c(2) = W - q(r)c(r) - q(g)c(g). \end{aligned}$$

Alternatively, we could derive the demand function for a single factor, holding other factor exposures constant. For example, consider the demand for the green factor holding fixed the exposure to the red asset. This corresponds to the conditional decision problem:

$$\begin{aligned} \max_{c(g)} \quad & \pi_2 u(c(2)) + \pi_1 \rho u(c(r)) + \pi_1(1-\rho)u(c(g)) \qquad \text{(Cond'l factor demand)} \\ \text{s.t.} \quad & c(2) = W - q(r)c(r) - q(g)c(g). \end{aligned}$$

We next derive the solution to these problems in terms of relative portfolio shares.

Proposition 4 (Zero factor elasticities) *Let $\omega^F(z) = q(z)c(z)/W$ denote the factor-level portfolio share. The conditional and unconditional relative portfolio shares satisfy*

$$\frac{\omega^F(z)}{\omega^F(2)} = \frac{\pi_z}{\pi_2} \qquad \textit{for } z \in \{r, g\}.$$

Hence factor demand elasticities are equal to zero for any factor prices.

Proof. The solution is standard. The first-order condition for Arrow security $z \in \{r, g\}$ is

$$\pi_2 q(z) u'(c(2)) = \pi_z u'(c(z)), \quad \text{that is,} \quad q(z) = \frac{\pi_z}{\pi_2} \frac{u'(c(z))}{u'(c(2))}.$$

Under log utility, this yields $q(z) = \frac{\pi_z}{\pi_2} \frac{c(2)}{c(z)}$. Imposing the budget constraint yields

$$c(z)q(z) = \frac{\pi_z}{\pi_2} \pi_2 W = \pi_z W, \quad \text{that is,} \quad \frac{c(z)q(z)}{W} = \pi_z.$$

Since portfolio shares are invariant in prices, the elasticity is always zero. ■

The result shows that factor demand functions exhibit low elasticities *for any prices and any payoff structures*. In particular, factor elasticities are zero for any value of ϵ. Since asset-level elasticities are strictly decreasing in ϵ and diverge to infinity as $\epsilon \to 0$, knowledge of the factor elasticities is not informative about the asset-level elasticities.

It is difficult to infer asset-level elasticities from factor elasticities mainly because no arbitrage ensures that factor prices are identical no matter through which asset factor exposures are acquired. Hence there is no asset-level variation in factors that can be used to identify cross-asset substitution.

Residual Demands We next measure residual demands in the case where factors do not subsume all asset cash flows. Because factors fully capture cash flows in our baseline model, we add idiosyncratic noise to green and red assets. The augmented payoff $\tilde{y}_j$ is

$$\tilde{y}_j = y_j + \eta_j,$$

where η_j is a random variable with mean μ_j, standard deviation σ_j that is uncorrelated across assets. Volatility σ_j is small in order to study a small perturbation of our model.

Next, suppose that the factors are directly traded, either because Arrow securities exist or because they are sufficiently many assets to form well-diversified factor portfolios. Let q_k denote the price of a unit of exposure to factor k and let each investor choose factor quantities α_k^i. An asset is a bundle of its factor exposures and the residual idiosyncratic component η_j. By no arbitrage, there exists a well-defined price for η_j, say $\tilde{p}_j$.

We can then model the portfolio choice in two steps. First, choose positions in the underlying assets. Second, adjust factor positions to achieve desired factor exposures. This leads to a natural *conditional* decision problem: controlling for factor exposures, choose positions $\tilde{a}_j$ for the idiosyncratic components of each asset. The solution to this problem yields *residual demand functions* given fixed factor positions. To make this point as simply as possible, assume as in Koijen and Yogo (2019) that the investor can invest in some outside asset in elastic supply, and that $\tilde{p}_j$ are again relative prices with respect to the outside good. Then the decision problem determining residual demand functions is

$$\max_{(\tilde{a}_j)} \quad \mathbb{E}[\log(\bar{c} + \sum_j \eta_j \tilde{a}_j)],$$

where

$$c = \underbrace{\sum_k \alpha_k F_k}_{\equiv \bar{c}} + \sum_j \eta_j \tilde{a}_j$$

is the consumption process for the investor and $\bar{c}$ is the (fixed) component of consumption that is due to factor exposures. As is standard, this problem can also be stated as:

$$\max_{(\tilde{\omega}_j)} \quad \mathbb{E}[\log(\bar{c} + W \sum_j \frac{\eta_j}{\tilde{p}_j} \tilde{\omega}_j)]$$

where $\tilde{\omega}_j = \tilde{p}_j \tilde{a}_j / W_j$ is the portfolio share of the residual component and we assume that changes in demand for the idiosyncratic demand is accommodated by a change in the demand for the outside asset. Since η_j is a payoff, $\eta_j / \tilde{p}_j$ is a return.

To illustrate the determination of residual demand elasticities, fix $\bar{c} = 1$ (as is the case in our model if $y(z) = 1$ and the residual component is small). Then a standard approximation to this problem as in Campbell-Viceira yields the decision problem:

$$\max_{(\tilde{\omega}_j)} \quad \mathbb{E}\Big[\sum_j \frac{\eta_j}{\tilde{p}_j} \tilde{\omega}_j\Big] - \frac{W}{2} \mathbb{V}\Big[\sum_j \frac{\eta_j}{\tilde{p}_j} \omega_j\Big].$$

Given that the idiosyncratic components are uncorrelated, optimal residual demand functions (relative to the outside good) have the simple form

$$\tilde{\omega}_j^* = \frac{\mu_j \, \tilde{p}_j}{\tilde{W} \, \sigma_j^2},$$

where σ_j is the variance of η_j. Observe that the residual elasticity is

$$\frac{\partial \tilde{\omega}_j^*}{\partial \tilde{p}_j} \frac{\tilde{p}_j}{\tilde{\omega}_j} = \frac{\mu_j}{\tilde{W} \, \sigma_j^2} \frac{\tilde{p}_j}{\tilde{\omega}_j} = 1.$$

This elasticity is low and constant, and uninformative about asset elasticities.

E Dynamic Trading

Consider a two-period variant of our baseline model. Trees are durable assets which pay dividends in two periods. Investors trade at the beginning of each period, and consume the per-period payoffs generated by their trees at the end of each period. The payoff structure is the same as in our baseline framework, and the aggregate state processes $z \in \{1, 2\}$ and $\iota \in \{g, r\}$, which determine payoffs, are *i.i.d.* across periods. The discount factor is given by $\delta \in (0, 1)$.

We denote by $\mathbf{S}_t$ the aggregate state variable sufficient for determining prices, which naturally includes the aggregate distribution over investor wealth and tastes. We use this structure to introduce potential variations in market prices over time. In particular, investors know the realization of $\mathbf{S}_1$ when forming portfolio allocations at date 1, and their choices are also influenced by their expectations of $\mathbf{S}_2$.

We write the price of asset $j \in \mathcal{J} \equiv \{g, r, 2\}$ as a function of the aggregate state: $p_{j,t} = P_j(\mathbf{S}_t)$ for some endogenous function P_j. The state-contingent gross return on asset $j \in \mathcal{J}$ is $R_j(\mathbf{S}_2) = \frac{P_j(\mathbf{S}_2)}{P_j(\mathbf{S}_1)}$. Given heterogeneity in tastes, a single investor will thus be concerned with the fact that changes in the preferences or wealth of other investors can induce changes in prices and, therefore, her perceptions of expected returns. It is sufficient for our purposes to consider the decision problem of a single investor who takes as given the stochastic process over the aggregate state. The investor's wealth at the beginning of period t is determined by the realized state and previous asset positions: $W_t^i(\mathbf{a}_{t-1}^i, \mathbf{S}_t) = \sum_{j \in \mathcal{J}} P_j(\mathbf{S}_t) a_{j,t-1}^i$.

We denote investor i's purchases of trees at the beginning of period $t \in \{1, 2\}$ by $\mathbf{a}_t^i = (a_{g,t}^i, a_{r,t}^i, a_{2,t}^i)$, and we denote by $\mathbf{a}_0^i$ investor i's exogenous endowment. The investor i's state variable at the beginning of period $t \in \{1, 2\}$ consists of the portfolio of asset positions purchased in the previous period, $\mathbf{a}_{t-1}^i$, and the current-period taste parameters $\Theta_t^i = (\theta_{j,t}^i)_{j \in \{g,r,2\}}$, which are permitted to evolve stochastically over time. The individual state at the beginning of period t is therefore $\mathbf{s}_t^i = (\mathbf{a}_{t-1}^i, \Theta_t^i)$.

We solve the problem by backwards induction, assuming that investors face short-sale constraints. For ease of exposition, assume that green and red trees are perfect substitutes: $\epsilon = 0$. In this case, the second-period choice between green and red trees is bang-bang: the investor buys only green trees if $\theta_{g,2}^i / P_g(\mathbf{S}_2) > \theta_{r,2}^i / P_r(\mathbf{S}_2)$, and only red trees if the inequality is reversed. We then have the following characterization of the second-period value function.

Lemma 2 (Value function) *Let $\epsilon = 0$ and $u = \log$. The second-period value function satisfies:*

$$V_2^i(\mathbf{s}_2^i, \mathbf{S}_2) = H(\Theta_2^i, \mathbf{S}_2) + \log\left(W_2^i(\mathbf{a}_1^i, \mathbf{S}_2)\right),$$

where $H(\Theta_2^i, \mathbf{S}_2)$ depends on investor i's tastes and market prices in period 2, but is independent of any investor choices at date 1.

Proof. Following the discussion of the static optimization problem, the second-period value function can be written as:

$$
\begin{aligned}
V_2^i(\mathbf{s}_2^i, \mathbf{S}_2) = \max_{\mathbf{a}_2^i \geq 0} \; \pi_1 \Big[\rho u\Big(\theta^i_{g,2} y_g(r) a^i_{g,2} + \theta^i_{r,2} y_r(r) a^i_{r,2}\Big) + (1-\rho) u\Big(\theta^i_{g,2} y_g(g) a^i_{g,2} + \theta^i_{r,2} y_r(g) a^i_{r,2}\Big)\Big] \\
+ \pi_2 u\left(\frac{y(2)}{P_2(\mathbf{S}_2)}\Big(W_2^i(\mathbf{a}_1^i, \mathbf{S}_2) - P_g(\mathbf{S}_2) a^i_{g,2} - P_r(\mathbf{S}_2) a^i_{r,2}\Big)\right).
\end{aligned}
$$

Since green and red assets are perfect substitutes (i.e., $\epsilon = 0$), the solution to the second-period decision problem is bang-bang, depending on whether $\theta_g^2 / P_g(\mathbf{S}_2) \geq \theta_r^2 / P_r(\mathbf{S}_2)$. Suppose that this inequality holds (i.e., green trees are cheap). Then, the value of the second-period problem is:

$$
V_{g,2}^i(\mathbf{s}_2^i, \mathbf{S}_2) = \max_{a^i_{g,2} \geq 0} \pi_1 u\Big(\theta^i_{g,2} y(1) a^i_{g,2}\Big) + u\left(\frac{y(2)}{P_2(\mathbf{S}_2)}\Big(W_2^i(\mathbf{a}_1^i, \mathbf{S}_2) - P_g(\mathbf{S}_2) a^i_{g,2}\Big)\right).
$$

The first-order condition together with log utility yields

$$
a^i_{g,2} = \frac{\pi_1 W_2^i(\mathbf{a}_1^i, \mathbf{S}_2)}{P_g(\mathbf{S}_2)}.
$$

This means that

$$
V_{g,2}^i(\mathbf{s}_2^i, \mathbf{S}_2) = \pi_1 u\Big(\theta^i_{g,2} \frac{\pi_1 y(1) W_2^i(\mathbf{a}_1^i, \mathbf{S}_2)}{P_g(\mathbf{S}_2)}\Big) + \pi_2 u\Big(\frac{y(2)}{P_2(\mathbf{S}_2)} \pi_2 W_2^i(\mathbf{a}_1^i, \mathbf{S}_2)\Big).
$$

With log utility this can be written as

$$
\begin{aligned}
V_{g,2}^i(\mathbf{s}_2^i, \mathbf{S}_2) = \pi_1 \log\left(\frac{\theta^i_{g,2}}{P_g(\mathbf{S}_2)}\right) + \pi_2 \log\left(\frac{1}{P_2(\mathbf{S}_2)}\right) + \pi_1 \log\Big(\pi_1 y(1)\Big) + \pi_2 \log\Big(\pi_2 y(2)\Big) \\
+ \log\Big(W_2^i(\mathbf{a}_1^i, \mathbf{S}_2)\Big).
\end{aligned}
$$

Similarly, if $\theta_g^2 / P_g(\mathbf{S}_2) \leq \theta_r^2 / P_r(\mathbf{S}_2)$, the value of the second-period problem is:

$$
\begin{aligned}
V_{r,2}^i(\mathbf{s}_2^i, \mathbf{S}_2) = \pi_1 \log\left(\frac{\theta^i_{r,2}}{P_r(\mathbf{S}_2)}\right) + \pi_2 \log\left(\frac{1}{P_2(\mathbf{S}_2)}\right) + \pi_1 \log\Big(\pi_1 y(1)\Big) + \pi_2 \log\Big(\pi_2 y(2)\Big) \\
+ \log\Big(W_2^i(\mathbf{a}_1^i, \mathbf{S}_2)\Big).
\end{aligned}
$$

Hence, the second-period value function is written as

$$V_2^i(\mathbf{s}_2^i, \mathbf{S}_2) = \max\left\{V_{g,2}^i(\mathbf{s}_2^i, \mathbf{S}_2), V_{r,2}^i(\mathbf{s}_2^i, \mathbf{S}_2)\right\},$$

and it satisfies

$$V_2^i(\mathbf{s}_2^i, \mathbf{S}_2) = H(\Theta_2^i, \mathbf{S}_2) + \log\Big(W_2^i(\mathbf{a}_1^i, \mathbf{S}_2)\Big),$$

where

$$\begin{aligned} H(\Theta_2^i, \mathbf{S}_2) =& \pi_1 \max\left\{\log\left(\frac{\theta_{r,2}^i}{P_r(\mathbf{S}_2)}\right), \log\left(\frac{\theta_{g,2}^i}{P_g(\mathbf{S}_2)}\right)\right\} + \pi_2 \log\left(\frac{1}{P_2(\mathbf{S}_2)}\right) \\ &+ \pi_1 \log\Big(\pi_1 y(1)\Big) + \pi_2 \log\Big(\pi_2 y(2)\Big). \end{aligned}$$

■

The separability of the value function into a taste component and a wealth component follows from log utility. However, this feature is not essential for our results below. What is important is that investors take into account future market prices when forming portfolios. Indeed, weakening separability would further complicate identification.

Now turn to the first-period decision problem. Maintaining the assumptions of log utility and normalizing $P_2(\mathbf{S}_1) = 1$, we can then write the period 1 decision problem under discount factor δ as:

$$\begin{aligned} V_1^i(\mathbf{s}_1^i, \mathbf{S}_1) = \max_{\mathbf{a}_1^i \geq 0} (1-\delta)\Bigg[&\pi_1 \log\Big(y(1)\Big(\theta_{g,1}^i a_{g,1}^i + \theta_{r,1}^i a_{r,1}^i\Big)\Big) \\ &+ \pi_2 \log\Bigg(y(2)\Big(W_1^i(\mathbf{a}_0^i, \mathbf{S}_1) - P_g(\mathbf{S}_1)a_{g,1}^i - P_r(\mathbf{S}_1)a_{r,1}^i\Big)\Bigg)\Bigg] \\ &+ \delta \mathbb{E}^i\Big[H(\Theta_2^i, \mathbf{S}_2) + \log\Big(\sum_{j \in \mathcal{J}} P_j(\mathbf{S}_2)a_{j,1}^i\Big)\Big], \end{aligned}$$

where the expectation $\mathbb{E}^i$ is taken with respect to private tastes and the aggregate state variable at date 2. Differentiating this object with respect to $a_{j,1}^i$ yields (13).

F Tastes versus constraints

We use the three-asset economy with the following simplifying assumptions:

(i) There are two types of investors, $i \in \{h, \ell\}$, each of which faces short-sale constraints. Tastes satisfy: $\theta_g^h = 1 + t$ and $\theta_r^h = 1 - t$, while $\theta_g^\ell = 1 - t$ and $\theta_r^\ell = 1 + t$.

(ii) Type h owns share $\gamma \geq \frac{1}{2}$ of the aggregate endowment of each tree.

(iii) There is no aggregate risk, $y(1) = y(2) = 1$ and $\pi_1 = \frac{1}{2}$.

We begin with the baseline where investors face only short-sale constraints. In this case, taste differences can lead to endogenous sorting in equilibrium.

First, look for an equilibrium in which type h specializes in green trees while type ℓ investor specializes in red trees. It follows from the first-order conditions with respect to a_g^h and a_r^ℓ and the aggregate resource constraint on tree 2 that

$$\frac{p_r}{p_g} = \frac{1 - a_2^h}{a_2^h} \qquad \text{and} \qquad p_1 \equiv \frac{p_g + p_r}{2} = \frac{\pi_1}{1 - \pi_1} = 1,$$

where the last equality follows from our simplifying assumptions. Substituting p_1 and the first-order condition with respect to a_g^h into type h investor's budget constraint, together with the aggregate resource constraint, one can show:

$$(a_2^h, a_2^\ell) = (\gamma, 1 - \gamma).$$

Thus, if sorting occurs in equilibrium, then the prices satisfy

$$(p_g, p_r) = \left(2\gamma \frac{\pi_1}{1 - \pi_1}, 2(1 - \gamma) \frac{\pi_1}{1 - \pi_1}\right) = (2\gamma, 2(1 - \gamma))$$

and the investors' portfolio choices are

$$(a_g^h, a_r^h, a_2^h) = \left(\frac{1}{2}, 0, \gamma\right) \qquad \text{and} \qquad (a_g^\ell, a_r^\ell, a_2^\ell) = \left(0, \frac{1}{2}, 1 - \gamma\right).$$

To show that this constitutes an equilibrium, we need to show that the first-order conditions with respect to a_r^h and a_g^ℓ hold at the zero holding. It can be seen that these first-order conditions are met as long as type h's incentive is satisfied:

$$\gamma \leq \overline{\gamma} \equiv \frac{\theta_g^h}{\theta_g^h + \frac{(y(1))^2 + \epsilon^2}{(y(1))^2 - \epsilon^2} \theta_r^h} = \frac{\theta_g^h}{\theta_g^h + \theta_r^h}.$$

Next, we consider an equilibrium in which one type of investor, denoted by i, holds both green and red trees while the other type specializes in one tree. Guessing that $\tilde{c}^i(g) = \tilde{c}^i(r) = \tilde{c}^i(2)$, first-order conditions with respect to a^i_g and a^i_r imply that

$$\left(p_g, p_r, p_1\right) = \left(\theta^i_g, \theta^i_r, 1\right).$$

Combining these conditions with the budget constraint implies that

$$a^i_2 = (1-\pi_1)e^i_2 + \pi_1 \frac{y(1)}{y(2)}(\theta^i_g e^i_g + \theta^i_r e^i_r) = \frac{e^i_2 + \theta^i_g e^i_g + \theta^i_r e^i_r}{2}.$$

We also guess and verify that $i = h$. Since type ℓ specializes in red, $a^\ell_g = 0$ and $a^h_g = \frac{1}{2}$. Then, by the aggregate resource constraint,

$$a^h_r = \frac{a^h_2 - \theta^h_g E_g}{\theta^h_r}.$$

Since the first-order condition with respect to a^ℓ_r yields $a^\ell_r = \frac{a^\ell_2}{p_r}$, it follows from the budget constraint that

$$a^\ell_2 = (1-\pi_1)(e^\ell_2 + p_g e^\ell_g + p_r e^\ell_r) = 1-\gamma \qquad \text{and} \qquad a^\ell_r = (1-\pi_1)\frac{e^\ell_2 + p_g e^\ell_g + p_r e^\ell_r}{p_r} = \frac{1-\gamma}{p_r}.$$

Thus, we obtain, as in the statement of the proposition,

$$a^\ell_r = \frac{1-\gamma}{2}\frac{E_2 + \theta^h_g E_g + \theta^h_r E_r}{\theta^h_r}.$$

When $\gamma > \overline{\gamma}$, it can be seen that $a^h_r > 0$ and that the first-order condition with respect to a^ℓ_g at $a^\ell_g = 0$ is also met (i.e., $a^\ell_g = 0$).

1 Introduction

How much do investors want to hold of a given asset, and how sensitive are their portfolio choices to the price? These questions lie at the heart of asset pricing. They determine how much prices move when central banks purchase bonds, when passive funds rebalance indices, or when financial intermediaries suffer shocks that force asset sales. An influential empirical literature has set out to answer them by estimating asset demand functions from data on portfolio holdings and prices.

A critical question for this literature is whether observed demand responses can be given a structural interpretation without relying on a fully specified equilibrium model of portfolio choice. If so, demand elasticities are invariant to the model used to estimate them and can be used to discriminate between different theories of investor behavior. If not, they are contingent, model-specific objects that reflect — rather than validate — a priori assumptions on investor behavior.

We provide a general theoretical analysis of this question. Our answer is that asset demand functions are not model-free empirical objects, and that structural modeling is unavoidable. Our results require only two foundational principles of asset pricing: that investors value assets for their payoffs, and that asset prices admit no arbitrage. As we discuss, these principles are difficult to discard without invalidating the basic premise of asset demand analysis.

Our results follow from a general decomposition of asset demand functions derived under preferences over payoffs and no arbitrage. Denote by Y the *payoff matrix* summarizing investor beliefs over the state-contingent payoffs of assets in the choice set. Then the matrix of asset demand slopes (in which each element is the derivative of asset demand with respect to a specific asset price) satisfies

$$\mathcal{A}^i = (Y^+)^{\mathrm{T}} \mathcal{D}^i Y^+.$$

In this decomposition, $\mathcal{D}^i$ is the investor's *fundamental demand function* for state-contingent payoffs and Y^+ is the Moore-Penrose pseudo-inverse of Y. This has a natural interpretation: since preferences are defined over payoffs, not assets directly, Y^+ maps demand for payoffs into the associated asset quantities.

This simple decomposition reveals two main challenges. First, the demand function for any individual asset is *commingled with those of all other assets*: because investors care about state-contingent consumption, the optimal quantity of any asset generically depends on the payoffs of all other assets through Y^+. This means that one cannot analyze demand for any given asset in isolation. Second, Y^+ is *latent and unidentifiable* from past data. Since the payoff matrix reflects investor beliefs about future payoffs, including resale prices, no finite sample of realized returns can pin it down—one can always alter the payoff of an unrealized state, changing Y^+ while leaving every historical return intact.

These features of the decomposition have immediate implications for asset demand analysis. Since Y^+ is unobservable, fundamental demand $\mathcal{D}^i$ cannot be recovered from portfolio data: different combinations of preferences and latent mappings are always observationally equivalent. And since Y^+ shifts whenever beliefs over payoffs are revised, asset demand functions are not structural with respect to standard perturbations that occur during regular market functioning.

What do these considerations imply for the estimation of asset demand functions without structural models? A common approach in the literature is to estimate *individual* asset demand curves through quasi-exogenous variation in asset supply. For this approach to work, supply shocks must generate *ceteris paribus* variation in a single asset price, holding all other prices and payoffs fixed.

Unfortunately, this identification assumption is generically inconsistent with the principle of no arbitrage and equilibrium price determination. Under minimal conditions, an increase in the supply of a given asset reduces the marginal cost of a unit payoff in a given state (i.e., the *state price*) in all states where the asset pays off. But by no arbitrage, this must lead to a change in the prices of all other assets that pay off in overlapping states. Since such payoff overlap is generic for essentially all asset markets, we prove that the identification of individual asset demand curves from individual supply shocks is generically infeasible. Most strikingly, we show that supply shocks generically imply state price changes that differ *directionally* from those required to estimate a fixed demand curve.

This leaves the possibility of jointly estimating the entire $J \times J$ *system* of demand slopes, where J is the number of assets, using multiple independent shocks to the price vector. This requires both at least J linearly independent price changes and that Y^+ remains fixed across all experiments. The first is a standard rank condition that arises in many settings; the second is the binding constraint. Since our decomposition shows that belief revisions generically shift all demand functions, it is implausible that the econometrician observes multiple independent shocks to the same demand system. Moreover, standard shocks used in the literature—such as central bank interventions or index inclusions—directly shift the demand system by altering future payoffs.

One might hope that imposing weak statistical structure on Y—for instance, through a factor model for asset returns—is enough to make progress. To investigate this, we use random matrix theory to study the statistical properties of the Moore-Penrose inverse Y^+ for factor-structured payoff matrices. Our results show the inverse mapping is generically ill-conditioned: the *sign* of any given element of Y^+ is a coin flip in large economies, and two economies sharing identical factor structures but different idiosyncratic payoff realizations have sign-independent inverses. Controlling for factor exposures therefore provides no systematic correction for the misalignment between supply shocks and the price variation required for demand estimation. We confirm these predictions numerically through Monte Carlo simulations and empirically using payoff data from S&P 500 stocks.

We summarize our results as a trilemma: given observational data, one cannot jointly maintain (i) no-arbitrage asset pricing, (ii) investor preferences over payoffs, and (iii) model-free identification of structural asset demand functions.

Our findings suggest a critical role for structural models in asset demand analysis. Since Y^+ cannot be identified from data, two models that agree on all observable implications of the data can imply arbitrarily different demand elasticities. For example, Fuchs, Fukuda, and Neuhann (2025) show that the logit model of Koijen and Yogo (2019) can infer an elasticity below one even if the true elasticity is infinite. Estimated asset demand elasticities should therefore not be treated as credible calibration targets, and should be evaluated on the plausibility and robustness of the assumed mapping rather than empirical fit.

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 that 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: investors form portfolios, marginal valuations depend on concurrent holdings of all other assets, the mapping from products to characteristics is latent, and choice is continuous. This fundamental non-separability of asset valuations under a latent mapping 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.

To overcome these challenges, structural models of asset demand must accurately account for the cross-asset linkages and price spillovers inherent to portfolio choice. Fuchs, Fukuda, and Neuhann (2025) show that the prominent logit approach in Koijen and Yogo (2019) can exhibit large biases in standard portfolio choice models with asymmetric substitution between assets. While our analysis in this paper focuses on contemporaneous cross-asset spillovers, similar 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. Perhaps most closely related to this paper is Haddad, He, Huebner, Kondor, and Loualiche (2025), who aim to recover relative demand elasticities from supply shocks without a structural model. Our findings suggest that their approach must impose theoretical restrictions if the estimated elasticities are to have a structural interpretation.

2 Framework

2.1 Environment

We study a canonical portfolio choice model. A mass of potentially heterogeneous investors I decide how much to consume at $t = 0$, and how to invest their savings to consume at $t = 1$. There are J financial assets, each of which yields a random payoff at date 1. Uncertainty is represented by a set of Z states of the world, one of which is realized at date 1. The payoff of asset j in state z is $y_j(z) \geq 0$. We denote by $Y \equiv (y_j(z))_{j,z}$ the $J \times Z$ matrix of cash flows. Since matrix Y reflects investors' beliefs about state-contingent payoffs, it is unobserved by the econometrician. We denote by $\pi \equiv (\pi_z)_z$, where $\pi_z \in (0, 1)$ is the probability of state z.

We are agnostic about the determinants of asset payoffs, and assume investors take the payoff matrix as given. However, in general the payoffs of a given asset are the sum of a direct cash component (i.e., dividends) and its expected resale price (i.e., the expected state-contingent market price). As we will discuss in more detail later, this means that one cannot easily assume that Y is a physical constant that remains fixed across time periods or economic regimes.

At time zero, each investor chooses a *portfolio* to maximize the expected utility of the state-contingent consumption across both dates. A portfolio is a vector of asset positions $a^i \equiv (a^i_j)^J_{j=1} \in \mathbb{R}^J$, where element a^i_j is the investor's holdings of asset j. Investor i's preferences are represented by a twice differentiable, strictly increasing and strictly concave von Neumann-Morgenstern utility function u^i.

Investors are competitive and take prices as given. The price of asset j is p_j, and time-zero consumption is the numeraire (or *outside asset*) with price normalized to one. Investor i is endowed with $e^i_j \geq 0$ units of asset j and $e^i_0 \geq 0$ units of the numeraire, and non-traded consumption endowments $w^i_0 \geq 0$ and $w^i(z) \geq 0$ at date 0 and in state z, respectively. Denote by $e^i \equiv (e^i_j)_j$ and $w^i \equiv (w^i(z))_z$. Portfolio choice may be curtailed by exogenous constraints: the investor must choose a portfolio from the set of feasible portfolios Φ^i, which we assume is a convex subset of $\mathbb{R}^J$.

Investor i's *portfolio choice problem* can then be formally stated as:

$$\begin{aligned} \sup_{a^i \in \Phi^i} \quad & (1-\delta^i)u^i(c^i_0) + \delta^i \pi \cdot u^i(c^i) \\ \text{s.t} \quad & c^i_0 = e^i_0 - p \cdot (a^i - e^i) + w^i_0 \quad \text{and} \\ & c^i = Y^{\mathrm{T}} a^i + w^i. \end{aligned} \tag{PCP}$$

Investor *i's asset span* $\mathcal{S}^i$ is set of payoff profiles that can be achieved through some feasible portfolio. That is,

$$\mathcal{S}^i \equiv \{Y^{\mathrm{T}} a^i \in \mathbb{R}^Z \mid a^i \in \Phi^i\}. \tag{1}$$

The portfolio choice problem embeds the canonical notion of *preferences over payoffs*: investors value state-contingent consumption and demand assets *instru-*

mentally for the payoffs they provide, not because they provide direct utility. Our results are robust to including a direct utility from holdings but are more sharply stated without them—all we require is that investors have at least some preferences over payoffs. A solution to this problem is jointly determined by several parameters: (i) the utility function u^i and rate of time preference δ^i, (ii) initial wealth w_0^i and state-contingent endowments w^i, which are demand shifters that shift state-specific marginal utility, (iii) portfolio constraints which determine the set of feasible portfolios Φ^i, and (iv) the payoff matrix Y and probability distribution π. We call the utility function, rate of time preference, demand shifters, and portfolio constraints *preference parameters*, which we denote by

$$\Theta^i \equiv \left(u^i, \delta^i, w_0^i, w^i, \Phi^i\right).$$

The optimal portfolio also depends on payoff matrix Y, which determines the mapping from asset positions to state-contingent payoffs, and the probability distribution π, which determines weights on states of the world. Since these objects do not pertain to investor preferences, we refer to these as *external parameters*.

Asset Demand Functions. A solution to problem (PCP) can be written in terms of J Marshallian *asset demand functions* which map parameters and the asset price vector into portfolio holdings. That is, the asset demand functional of investor i is

$$a^i(\cdot \mid \Theta^i, Y, \pi) : \mathbb{R}^J_{++} \to \mathbb{R}^J.$$

Standard portfolio choice theory shows that all asset demand functions generically depend on the entire vector of asset prices. That is, asset demand is *non-separable*.

In line with empirical practice, we will typically focus on identifying the $J \times J$ matrix of asset demand derivatives given a prevailing asset price vector p:

$$\mathcal{A}^i(\Theta^i, Y, \pi) \equiv -\frac{\partial a^i(p \mid \Theta^i, Y, \pi)}{\partial p^{\mathrm{T}}}.$$

This object characterizes asset demand within a neighborhood of price vector p.

In general, the econometrician observes neither preference parameters Θ^i nor external parameters (Y, π). The demand identification problem thus is to infer combinations of these parameters which determine asset-level demand functions and are invariant to perturbations or counterfactuals of interest.

2.2 Consistent Pricing Systems and No Arbitrage

Before analyzing the demand identification problem in more detail, we establish the importance of a consistent pricing system for all possible portfolios of assets. This motivates our use of no arbitrage to structure the pricing system.

Necessity of internally consistent prices. A defining feature of portfolio choice is that investors can flexibly bundle and unbundle assets to achieve desired payoff processes. For example, two assets with state-contingent payoffs $[1, 1]$ and $[1, 0]$ can be combined into a portfolio with payoff $[0, 1]$, or indeed *any* payoff in $\mathbb{R}^2$. Given continuous choice over assets, investors can thus choose among a *continuum* of potential payoff vectors whose mapping into portfolios depends on the unobserved payoff matrix Y. To permit inference about preferences from portfolio holdings, the econometrician must therefore impose a priori assumptions on the pricing system which can be used to construct prices for all feasible payoff vectors.

No arbitrage ensures consistent pricing. The canonical approach to ensuring consistent pricing in financial markets is the principle of *no arbitrage*, which asserts that pricing system should not permit trading strategies which offer "something for nothing." In particular, this principle states that there should not exist feasible trading strategies which offer strictly positive payoff at some date while offering weakly positive payoffs in all other states and dates.

Definition 1 (No Arbitrage) *There is no arbitrage if there does not exist a portfolio $a^* \in \mathbb{R}^J$ such that $Y^{\mathrm{T}} a^* \geq 0$ and either (i) $p \cdot a^* \leq 0$ and$(Y^{\mathrm{T}} a^*)_z > 0$ for some z or (ii) $p \cdot a^* < 0$.*

No arbitrage is a weak restriction which rules out the existence of profitable trading strategies that would be exploited by any investor with increasing preferences. Nevertheless, it is sufficient to ensure consistent pricing of *all* portfolios. In particular, the fundamental theorem of asset pricing shows that no arbitrage is equivalent to the existence of a vector of *state prices* $q \in \mathbb{R}^Z$ which serve as reference prices from which one can recover any asset price. State prices can be interpreted as the marginal cost of unit payoff in a given state of the world, so that asset prices are payoff-weighted sums of state prices. See Duffie (2001) for the proof.

Theorem 0 (Fundamental Theorem of Asset Pricing) *Let $\Phi^i = \mathbb{R}^J$. There is no arbitrage if and only if there exist state prices $q \in \mathbb{R}^Z_{++}$ such that asset prices satisfy*

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

Under no arbitrage, the prices of all traded portfolios are thus linear combinations of state prices, with weights determined by state-contingent payoffs. This yields a consistent pricing system that links asset prices to the marginal cost of what they offer to investors, namely state-contingent payoffs. Because of its foundational role in modern asset pricing, we adopt this pricing system as well.

Existence of Optimal Portfolios and Dimension Reduction. No arbitrage provides two additional advantages for asset demand analysis. First, it allows researchers to combine individual asset positions into aggregated portfolios while ensuring that the demand for and price of the bundle is internally consistent. Such portfolio aggregation is a foundational tool. For example, Koijen and Yogo (2019) aim to summarize asset demand using a small number of asset characteristics. Second, a canonical result—recapitulated below—shows that no arbitrage ensures the *existence* of a solution to the portfolio choice problem. Naturally, existence is a prerequisite for demand analysis in asset markets. See Duffie (2001) for the proof.

Proposition 0 (No arbitrage and the investor's problem) *Let $\Phi^i = \mathbb{R}^J$. Then there is a solution to (PCP) if and only if there is no arbitrage.*

Taken together, the principle of no arbitrage thus serves to ensure the internal consistency of asset demand systems while imposing only weak assumptions. While some trading frictions could prevent no arbitrage from holding exactly, as long as the frictions do not completely rule out general equilibrium price adjustments, our arguments hold.

Redundant assets. No arbitrage pricing is particularly salient in the presence of redundant assets (i.e., when there are multiple portfolios that deliver identical cash flow processes). In such cases, an arbitrarily small change in the price of a redundant asset immediately triggers an arbitrage opportunity with discontinuous changes in demand functions (see Example 3 in Appendix D.1 for an illustration). If such arbitrages do persist on the equilibrium path, it is infeasible to identify asset-specific demand functions (i.e., the slope of asset quantities with respect to variation in a single price) for redundant assets from observational data. For the remainder, we therefore focus on the case without redundant assets.

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

2.3 Outline of the Argument

Our argument proceeds in three steps. *First* (Section 3), we derive the decomposition $\mathcal{A}^i = (Y^+)^\mathrm{T}\mathcal{D}^i Y^+$, which separates asset demand slopes into fundamental preferences over payoffs and a latent mapping Y^+ from payoff demand into portfolios. Since Y^+ is unobservable in principle, asset demand functions are not structural with respect to the belief revisions that occur during ordinary market functioning.

Second (Section 4.1), we ask whether the latent mapping can be bypassed through supply shocks and show that it cannot: no-arbitrage forces prices of all payoff-overlapping assets to move jointly, so supply shocks generically produce price variation that is misaligned with the ceteris paribus requirements of demand identification.

Third (Section 4.3), we ask whether imposing factor structure is sufficient to make progress and show that it is not: using random matrix theory, we establish that Y^+ is generically ill-conditioned, with individual elements having the wrong sign with probability approaching one-half in large economies. Section 5 collects these findings into the trilemma and draws implications for the interpretation of estimated demand elasticities.

3 The Problem of the Unobservable Mapping

To understand basic properties of asset demand function, we begin by establishing a general decomposition of asset demand functions into two components: fundamental demand over state-contingent consumption, and a *latent mapping* which determines the asset portfolio required to achieve target state-contingent consumption profile. The first component reflects the standard notion of demand that is common to all demand analysis, in that it reflects preference parameters and willingness to pay for consumption in different states of the world. The latent mapping is unique to the case of financial assets, in that it reflects beliefs over future payoffs which guide how different assets must be combined with each other.

Our main result is that the latent mapping is fundamentally *unobservable*. This has two implications: (i) fundamental demand can never be identified from data on portfolio choices, and (ii) asset demand can be given a structural interpretation (that is, being invariant to perturbations) only if the generalized inverse of the payoff matrix Y remains fixed. Since payoff and forecast revisions occur across essentially all time horizons and financial markets, this suggests that asset demand is not a structural object in essentially all settings of interest.

3.1 Demand Decomposition and Non-Separability

To arrive at our decomposition, we must define an appropriate notion of demand functions for state-contingent consumption. We thus consider the following *con-*

sumption choice problem given a vector of state prices.

$$\max_{c^i \in \mathcal{C}^i} \quad (1-\delta^i)u(c_0^i) + \delta^i \pi \cdot u(c^i) \tag{CCP}$$
$$\text{s.t} \quad c_0^i + q \cdot (c^i - w^i) \leq w_0^i + q \cdot Y^{\mathrm{T}} e^i.$$

A solution to problem (CCP) is a set of Z state-contingent consumption functions c^i which map the vector of state prices into a consumption profile. The budget constraint states that net purchases of consumption (over and above endowments) must be equal to the value of the investor's asset endowments. The set $\mathcal{C}^i$ encodes constraints on feasible consumption choices due to, e.g., incomplete markets or other portfolio constraints. By analogy with asset demand functions, we can therefore study the derivative of the Marshallian *consumption* demand curve with respect to *state prices*.

Definition 2 (Consumption demand) *The slope of consumption demand is a $Z \times Z$ matrix of Marshallian consumption demand slopes with respect to state prices,*

$$\mathcal{D}^i \equiv -\frac{\partial c^i}{\partial q^{\mathrm{T}}}$$

which depends on preference parameters, state probabilities π and state prices but is independent of the payoff matrix conditional on investor i's asset span $\mathcal{S}^i$ defined by (1).

Consumption demand functions thus have a certain robustness property with respect to small perturbations of the payoff matrix Y. When these perturbations do not affect the set of attainable payoffs, they do not alter consumption plans. As we will see, this is *not* the case for asset demand functions, which do depend directly on perturbations of Y.

However, consumption demand functions are not observable because neither state prices nor the payoff process are directly observed by the econometrician. However, they can be linked to observable portfolios and asset prices using the portfolio choice problem and no arbitrage pricing. This follows directly from the chain rule. Specifically, the consumption process induced by portfolio is $c^i = Y^{\mathrm{T}} a^i + w^i$. The portfolio yielding a target consumption process c^* thus is $a^*(c^*) = (Y^+)^{\mathrm{T}}(c^* - w^i)$, where Y^+ is the generalized Moore-Penrose inverse of Y. Differentiating asset demand with respect to asset prices therefore yields

$$\frac{\partial a^*(c^*)}{\partial p^{\mathrm{T}}} = (Y^+)^{\mathrm{T}} \frac{\partial c^*}{\partial p^{\mathrm{T}}} = (Y^+)^{\mathrm{T}} \frac{\partial c^*}{\partial q^{\mathrm{T}}} \frac{\partial q^{\mathrm{T}}}{\partial p^{\mathrm{T}}},$$

where the second equality follows from the chain rule. Next, observe that no arbitrage implies that asset prices p are related to state prices through $p = Yq$, and thus state prices can be related to asset prices through $q = Y^+ p$. Applying this observation to $\frac{\partial q^{\mathrm{T}}}{\partial p^{\mathrm{T}}}$ then yields the following decomposition.

Proposition 1 (Demand Decomposition) *If asset prices satisfy no arbitrage, then*

$$\mathcal{A}^i = (Y^+)^{\mathrm{T}} \mathcal{D}^i Y^+. \tag{3}$$

If preferences over state-contingent consumption are separable across states as in (PCP) *and there are no binding portfolio constraints, then one can further decompose*

$$\mathcal{A}^i = (Y^+)^{\mathrm{T}} \Pi^{-1} \tilde{\mathcal{D}}^i Y^+, \tag{4}$$

where Π is a diagonal matrix of state probabilities and $\tilde{\mathcal{D}}^i$ depends only on preferences.

The decomposition reveals that asset demand is generically *non-separable*: given a desired consumption process, the optimal position in any given asset is jointly determined by the payoff processes of *all* assets in the choice set. In particular, changing the payoff characteristics (state-contingent payoffs) of any given asset generically alters several elements of the inverse payoff matrix, triggering changes in the optimal demand for other assets.

A particularly stark illustration of this fact is that whether two assets are substitutes or complements depends on the attributes of other assets in the choice sets. To the best of our knowledge, this issue is distinct from other settings in industrial organization, which may consider flexible specifications in which two goods can be substitutes or complements, but these parameters are invariant to attributes of other goods. This however is a central feature of the portfolio choice.

We illustrate this with a simple 3 asset example, which satisfies the condition that any perturbations of the payoff structure do not change the asset span.

Example 1 (Complementarity and substitutability depend on other assets) *Consider a three-asset, three-state economy with the payoff matrix Y, where rows index assets:*

$$Y = \begin{bmatrix} 1 & \epsilon & 0 \\ \epsilon & 1 & 0 \\ \xi & \xi & 1 \end{bmatrix},$$

where $\xi, \epsilon \in (0,1)$. Thus, we have complete markets. Assuming homogeneous consumption elasticities, the cross-elasticity between Assets 1 and 2 is:

$$-\frac{\partial a_1}{\partial p_2} = \sum_z (Y^{-1})_{z,1}(Y^{-1})_{z,2} = \frac{\xi^2(1-\epsilon)^2 - 2\epsilon}{(1-\epsilon^2)^2}.$$

For a given ϵ we can find the value of ξ that sets the cross-elasticity to zero:

$$\xi^* = \frac{\sqrt{2\epsilon}}{1-\epsilon}.$$

Thus Assets 1 and 2 complements for $\xi < \xi^*$ and substitutes for $\xi > \xi^*$.

Figure 1 illustrates the change in complementarity as a function of ξ. To understand the intuition for this example, consider first a very low value of ξ. In this case, Asset 3 plays an insignificant role in the payoff of the first two states and Assets 1 and 2 are good hedges for each other in states 1 or 2. This force is stronger when ϵ is smaller. Next, consider a very large value of ξ: Asset 3 now replicates the payoffs of Assets 1 and 2 in states 1 and 2 ever more closely, crowding out the hedging roles of both. When the price of Asset 2 rises, investors turn to Asset 3 as an alternative hedge for state 2, but in doing so they simultaneously acquire exposure to state 1. This reduces their demand for Asset 1 as well. The two assets thus become substitutes not through any direct relationship between them, but because they share a common "competitor." This force is stronger when ϵ is larger.

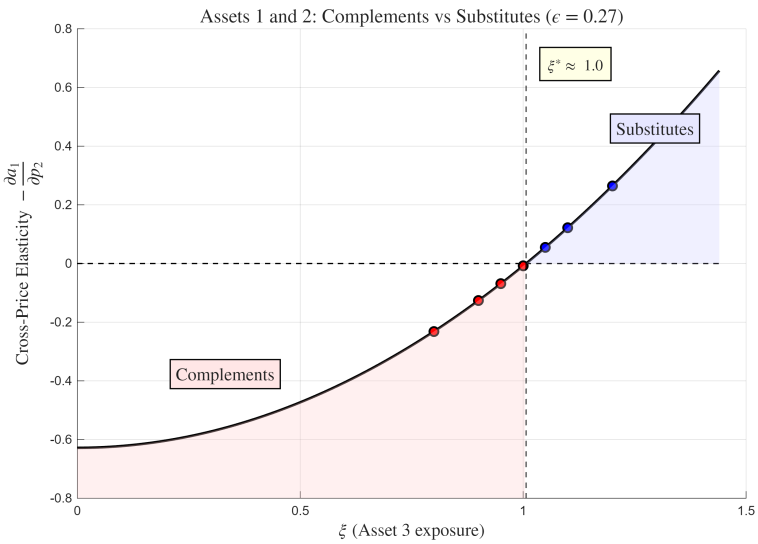


Figure 1: Switch in complementarity given ξ in Example 1

3.2 Unobservable Mapping

We have shown that asset demand functions—as well as the degree of complementarity between any two assets—is determined by global properties of the inverse payoff matrix Y^+. We next show a fundamental constraint on asset demand analysis: the mapping from fundamental preferences to portfolio choices is unobservable because the generalized inverse payoff matrix Y^+ cannot be identified from any finite sample of observed payoffs *even if the payoff matrix is assumed to be stable.*

Proposition 2 (Non-identification of the Latent Mapping Y^+) *Consider any finite sample of realized asset payoffs $\mathcal{S}$. Then there exist arbitrarily many candidate payoff matrices Y which are observationally equivalent given $\mathcal{S}$ but have different generalized inverses.*

Intuitively, the proposition holds because a researcher can always alter the payoff of a state of the world that has not yet been realized in the data, and altering payoffs in this state changes Y^+ but leaves all observed historical returns identical. This fundamental non-identification of Y^+ implies that asset demand estimation must reckon with two commingled identification problems: that of fundamental preferences (i.e., the standard identification problem that is common to all demand estimation exercises), and that of the latent mapping linking the primitive object of preferences (payoffs) to observed choices (asset positions).

An important implication of this is that fundamental demand can never be identified from observed portfolio choices. The reason is that observed choices and preferences pertain to different objects, namely assets and payoffs, and the mapping between the two is unobservable. As such, observed asset choices can always be rationalized by different combinations of preferences and the latent mapping.

Corollary 1 (Non-recoverability of $\mathcal{D}^i$) *For any $\mathcal{A}^i$, the consumption demand slope $\mathcal{D}^i$ is generically not uniquely determined absent knowledge of Y^+.*

Proof. The result follows because Y^+ cannot be identified from data (Proposition 2) and different inverses Y^+ generically induce different $\mathcal{D}^i$ through equation (3) even for the same $\mathcal{A}^i$. ■

3.3 Asset Demand is Not Structural under Weak Conditions

We now use our decomposition to provide conditions under which demand functions are *structural* in the sense of Hurwicz (1962) and Marschak (1953): invariant to relevant perturbations to the economic environment. To formalize this, we allow all model primitives to vary with a latent variable $\omega \in \Omega$ which we call the *economic environment*. We then say that a demand function is *structural* with respect to a class of perturbations if it is invariant across all ω within the class.

Definition 3 (Structural Demand) *Let $\mathcal{P} \subseteq \Omega \times \Omega$ be a class of perturbations, where each $(\omega, \omega') \in \mathcal{P}$ represents a transition from environment ω to ω'. We say that a demand function $\mathcal{F}$ is structural with respect to $\mathcal{P}$ if:*

$$\mathcal{F}(\omega) = \mathcal{F}(\omega') \quad \textit{for all } (\omega, \omega') \in \mathcal{P}.$$

The structural properties of demand thus depend on the perturbations under consideration. While these are often application specific, we emphasize perturbations that are of particular interest to financial markets: *unobserved* revisions to payoff expectations Y and state probabilities π which occur in the regular course of financial market operations. If asset demand is not structural with respect to these perturbations, there can be no guarantee that demand is structural with respect to other interventions or perturbations either.

Our decomposition then implies two main results: consumption and demand functions cannot be jointly be structural with respect to perturbations that vary the payoff matrix, and asset demand functions can be structural only if consumption preferences are a specific function of payoff parameters themselves. This contradicts the fundamental dichotomy between investor preferences and asset characteristics which permits demand analysis in the first place.

Proposition 3 (Non-structural demand) *Consider a class of perturbations with unobservable changes to Y or π. Then:*

1. *Asset demand function $\mathcal{A}^i$ and $\mathcal{D}^i$ cannot both be structural.*
2. *If asset demand functions are to be structural, then fundamental preferences must respond to any shock to future prices, dividends, or probabilities. In particular, we must have $\mathcal{D}^i = Y^T B^i Y$ for an arbitrary B^i, so that $\mathcal{A}^i = B^i$.*

That is, one cannot maintain the assumption that demand functions are structural unless one is also willing to make the assumption that probabilities and payoffs must remain fixed. This presents a sharp constraint on asset demand analysis because the assumption is (i) unverifiable, and (ii) it rules out that asset demand is structural with respect to interventions whose *goal* is to shift payoff expectations. This includes quantity-based policy experiments such as quantitative easing or foreign exchange interventions which aim to influence broader economic conditions. Such applications are of central interest to much of the literature.

4 The Problem of Identification

We now establish conditions under which asset demand can be identified from observational data on portfolio holdings and asset prices. We provide two main results. First, we show that the canonical approach to estimating demand curves—namely, exogenous shocks to the supply of a given asset—does not identify individual demand curves because they generically fail to produce ceteris paribus variation in a single asset price. Second, we ask whether multiple supply shocks can be combined to identify the $J \times J$ matrix of demand slopes $\mathcal{A}^i$. We show this to be possible if (i) the econometrician observes at least J linearly independent, exogenous shocks to the price vector and associated portfolio responses and (ii) the unverifiable assumption that the latent mapping Y^+ remains fixed across all such experiments. While the requirement of J shocks is a standard rank condition that also arises other settings with demand complementarities, the second is specific to financial markets, where the mapping between assets and characteristics is unobservable. Since forecast revisions are a defining function of financial market, this assumption is both strong and unverifiable in essentially all settings of interest. Furthermore, there is a natural tension: many experiments which shift prices today are likely to shift future prices as well, but this would also change Y^+, which contains resale prices.

4.1 Supply Shocks Produce Misaligned Price Variation

The canonical approach to estimating demand curves is to rely on supply shocks to provide suitably exogenous variation in a given price. With demand complementarities and long-lived assets, a central endogeneity concern is that the supply shock creates correlated changes in other asset prices (so-called *price spillovers*), thereby contaminating the price variation needed to identify a particular demand slope. We now show that this problem is generic under no arbitrage: except in the implausible knife-edge case where assets never pay off in overlapping states of the world, even perfectly exogenous supply shocks must always induce price spillovers.

Ideal experiment. We first define the price variation necessary to identify a particular asset demand slope. Given that asset demand exhibits demand complementarities and depends on expected future resale prices, we require that all other prices and all future payoffs must remain unchanged. To understand whether such variation is obtainable under even ideal conditions, we study the price changes induced by *perfectly exogenous supply shocks that leave all future payoffs unchanged.*

Under preferences over payoffs, it is useful to describe the ideal experiment in terms of state prices. The investor observes asset prices p and payoff matrix Y. Under no arbitrage, prevailing asset prices imply the state price vector

$$q = Y^{+} p, \tag{5}$$

where Y^{+} is the Moore-Penrose pseudo-inverse of Y.[1] A hypothetical *pure price shock* to asset j thus induces a specific state price change which is fully determined by Y^{+}.

Lemma 1 (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*

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

The assertion follows immediately from equation (5). Identifying asset demand thus requires shocks which generate the state price variation $\Delta \mathbf{q}_j^{\text{ideal}}$ associated with the ideal experiment.

Measurement using supply shocks. Since pure price shocks are rarely observed, we now study whether even perfectly exogenous supply shocks can generate the variation required by the ideal experiment. 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

[1] 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 exist multiple feasible state price vectors. As is standard, we select the minimum norm solution with pseudo-inverse $Y^{+} = Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}$, which also arises endogenously in our demand decomposition.

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 call this property *downward-sloping consumption demand*. Since our definition is written directly in terms of state prices, it must be understood purely in terms of fundamental preference parameters.

Definition 4 (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*

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

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 4 imposes no assumptions on V's off-diagonal entries, which capture potential *direct* preference-based spillovers across state prices in response to a supply shock. The existence of such spillovers depends on the economic model. The canonical model with additively separable utility over consumption (as in Section 2) has zero off-diagonal elements. Example 2 below illustrates this with a representative investor. Non-separable models such as recursive utility (Epstein and Zin, 1989; Kreps and Porteus, 1978) or more general aggregators instead generally imply non-zero off-diagonal elements. Since spillovers are the main threat to identification, the identification challenge is generically *weaker* when there are no preference-based spillovers in state prices. To provide favorable conditions for identification, we thus assume that no such spillovers exist.[2]

[2]The only case in which non-diagonal V can undo price spillovers occurs when the off-diagonal elements in V exactly offset the cross-asset restrictions implied by no arbitrage. However, V is determined by preferences and aggregate endowments while the no-arbitrage relation depends only on the payoff matrix Y. Hence there is no economic reason for such a mechanical offset to occur. More generally, Section 4.3 shows that, for large matrices, 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 meaningfully altering V.

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.*

Example 2 (V in an additive separable representative-agent model) *In a standard representative-agent model with additive separable preferences over consumption, state 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 pricing kernel is a strictly positive diagonal matrix,

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

Supply Shocks Do Not Generate the Ideal Experiment. We now show that supply shocks generically fail to produce the ideal experiment. We consider two definitions of alignment between supply shocks and the ideal experiment: (i) that induced state price changes are identical to those of the ideal experiment (up to a scalar multiple), and (ii) that the induced state price changes are of the same *sign*.

While the first condition is required to exactly identify a demand slope, the second captures the much weaker requirement that the supply shock should at least trigger *directionally consistent* changes in the cost of consumption. If this condition 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 lead to large biases when estimating substitution patterns.

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

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.*

We can then state our main result of this section, which is 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 that state. That is, in order to satisfy the minimal requirement that the induced state price variation is of the same sign as in the ideal experiment, there must be no assets with overlapping payoffs.

Definition 5 (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 (Supply Shocks Induce Misaligned Price Variation) *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.

The conditions set out 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 where overlap in payoffs (that is, concurrent non-zero dividends and/or resale values) is the norm, not the exception. It is therefore striking that, outside of these knife-edge restrictions, supply shocks do not even guarantee *directional* alignment with the ideal experiment. As such, supply shocks generically fail to identify structural asset-level demand slopes in essentially all settings of interest. In Appendix B, we also illustrate our findings using a simple example economy based on Fuchs, Fukuda, and Neuhann (2025).

Asset-by-asset misalignment. Theorem 1 shows that there must be misaligned price variation for at least one asset in the payoff menu (that is, at least one row of the payoff matrix). This allows the possibility that misalignment may not occur for *some* assets (although this cannot be verified without knowledge of the payoff matrix). The next proposition further strengthens our result by providing a weak condition for which misaligned price variation is guaranteed for *every asset.*

Proposition 4 *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 $(Y^+)_{z,j} < 0$.*

4.2 Identification from Multiple Shocks Requires Fixed Payoffs

So far we have shown that even perfectly exogenous supply shocks generically produce price variation that is contaminated by cross-asset spillovers. While this means that one cannot identify specific asset demand curves from asset-level supply shocks, one might yet jointly identify the entire matrix $\mathcal{A}^i$ by combining sufficiently many independent supply shocks.

We now show that this is the case only if (i) the econometrician has access to at least J exogenous supply shocks which provide linearly independent variation in the price vector—a standard rank condition—and (ii) that the unobservable mapping Y^+ remains fixed across all supply shocks. The second condition is a central challenge in financial markets, where the mapping from goods to characteristics is unobserved and subject to frequent revisions over essentially any time horizon.

Setting. We consider an idealized scenario in which the econometrician observes K independent shocks to the price vector and interprets investors' portfolio responses under a set of maintained assumptions $\mathcal{M}$. We refer to each price shock as an *experiment*, and assume that they are linearly independent. Since there are J assets, it is sufficient for our argument to consider the case where the number of experiments is weakly smaller than the number of assets: $K \leq J$. Let $\mathcal{O}_P$ denote the observed price changes, and $\mathcal{O}_{a^i}$ the observed portfolio response for investor i. We have the following standard definition of identification.

Definition 6 (Identified Demand Functions) *Two asset demand functions $a^i(\cdot \mid \Theta^i, Y, \pi)$ and $a^i(\cdot \mid \tilde{\Theta}^i, \tilde{Y}, \tilde{\pi})$ are observationally equivalent given data (O_P, O_{a^i}) and maintained* *assumptions $\mathcal{M}$ if both satisfy $\mathcal{M}$ and, for each observed price vector $p \in O_P$,*

$$a^i(p \mid \Theta^i, Y, \pi) = a^i(p \mid \tilde{\Theta}^i, \tilde{Y}, \tilde{\pi}) = O_{a^i}.$$

Demand slope $\mathcal{A}^i$ is identified under maintained assumptions $\mathcal{M}$ if all observationally equivalent demand functions imply the same slope:

$$\mathcal{A}^i(\Theta^i, Y, \pi) = \mathcal{A}^i(\tilde{\Theta}^i, \tilde{Y}, \tilde{\pi}).$$

We have already established that $\mathcal{D}^i$ is not identified because Y^+ is not observable. We now show that asset demand is identifiable only if the econometrician observes J experiments *and* maintains the unverifiable assumption that the unobservable mapping Y^+ is fixed across all experiments.

Proposition 5 (Necessary Conditions for Identification of $\mathcal{A}^i$) *Let (O_P, O_{a^i}) denote a data set of K observed price vectors and associated portfolio positions for investor i. Generically, $\mathcal{A}^i$ is identified only if the econometrician observes $K = J$ independent experiments and maintains the assumption that Y^+ is fixed across all experiments.*

Identification thus necessarily relies on the unverifiable assumption that the unobservable inverse payoff matrix is fixed across at least J shocks to the price vector. This finding relates our work to Haddad, He, Huebner, Kondor, and Loualiche (2025), who aim to recover an asset elasticity matrix without a fully specified structural model by combining (i) cross-sectional variation in asset-level holdings, and (ii) time series shocks to the prices of certain factor portfolios. Proposition 5 shows that this approach identifies a well-defined and stable elasticity matrix only if the unobserved payoff matrix remains fixed over time. However, changes in expected payoffs are a natural byproduct of financial market activity, including risk sharing, investment, or price discovery. (Section 4.3 shows that even approximate stability in the payoff matrix does not ensure a stable *inverse* payoff matrix, which is what matters for the stability of asset demand functions.)

Implications for Instrument Validity. Propostion 5 also sharpens the conditions for instrument validity in asset demand estimation: in addition to providing suitably exogenous variation in prices *today*, the instrument must *not* alter future expected payoffs. This rules out any shock to current prices that simultaneously induces changes in expected asset payouts or resale prices, such as central bank asset purchases or index inclusion events. The former are often used precisely to influence expectations over future market conditions, whereas index inclusion is known to alter return comovements and the level of prices.

4.3 The Latent Mapping is Ill-conditioned and Unstable

We have shown that asset demand functions are neither structural nor identifiable unless the latent inverse payoff matrix is assumed to remain fixed. These issues arise because one can at best observe a subset of *realized* payoffs, but not the matrix of *expected* payoffs, although it is the latter which matters for asset demand.

A potential solution to this problem is to use statistical information on realized returns to impose structure on the payoff process, and to hope that this structure is sufficient to ensure a stable mapping from preferences to asset holdings. The predominant approach in the literature is to impose a factor structure on payoffs, whereby asset returns are driven by a relatively small number of common factors. We therefore use random matrix theory to analyze the asymptotic properties of factor-structured payoff processes.

We find that the inverse payoff matrix is ill conditioned: the *sign* of any given element of the inverse payoff matrix is a coin flip. Hence even well-behaved factor structures yield poorly behaved latent mappings that can flip signs even with small changes to the payoff process. Monte Carlo simulations show that our theoretical limit results hold even for small J and Z. Appendix C shows that similar results hold in data from the S&P500. We conclude that imposing realistic statistical structure on the payoff matrix is *not* sufficient to ensure a well-behaved mapping from preferences to asset positions.

Random matrix approach. 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:

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

The analysis extends to multi-factor processes: see Remark 2 in Appendix A.4. As before, let Y^+ denote the Moore-Penrose pseudo-inverse of Y.[3] We impose the following assumptions.

Assumption 3 (Factor structure) *$(\alpha_j, \beta_j)_j$ are i.i.d., independent of $(f_z)_z$ and $(\varepsilon_{j,z})_{j,z}$, with finite second moments.*

Assumption 4 (Factor returns) *$(f_z - \overline{f})_z$ are i.i.d. with bounded, continuous, and symmetric densities around 0, and $\sigma_f^2 \equiv \mathbb{V}[f_z] < \infty$.*

Assumption 5 (Idiosyncratic shocks) *$(\varepsilon_{j,z})_{j,z}$ are i.i.d. across (j, z) with bounded, continuous, and symmetric densities around 0, and $\sigma_\varepsilon^2 \equiv \mathbb{V}[\varepsilon_{j,z}] > 0$. Factors and errors are mutually independent.*

Population objects and sequential limits. The properties of small random matrices are difficult to characterize with any generality. Theorem 2 thus considers the sequential limit $Z \to \infty$ followed by $J \to \infty$.[4] As $Z \to \infty$ the sample Gram matrix $G_Z \equiv \frac{1}{Z} Y Y^{\mathrm{T}}$ converges almost surely to the population second-moment matrix,

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

[3] The rank of Y equals J almost surely under Assumption 5, since the set of $J \times Z$ matrices with $\text{rank}(Y) < J$ has measure zero. Hence $Y^+ = Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}$ a.s.

[4] The sequential limit $Z \to \infty$ followed by $J \to \infty$ is adopted for transparency of proof, not out of necessity. Our numerical simulations suggest that Z and J could be taken to infinity at the same time yet allowing for that would significantly complicate the proof.

The sign of $(Y^+)_{z,j}$ is asymptotically determined by the sign of the population quantity $(\Sigma^{-1}y_z)_j$, which can also be written as $v_j^\top \Sigma^{-1} y_z$ where $v_j \in \mathbb{R}^J$ is the j-th unit vector. We thus work directly with the following population objects:

1. *Individual sign probability.* For each fixed (j, z), define

$$\pi(j,z) \equiv \lim_{Z\to\infty} P\left((Y^+)_{z,j} > 0\right).$$

2. *Sign-agreement frequency.* Let Y and $\tilde{Y}$ be two payoff matrices generated by the same factor loadings $(\alpha_j, \beta_j)_j$ and factor realizations $(f_z)_z$ but independent idiosyncratic shocks $(\varepsilon_{j,z})$ and $(\tilde{\varepsilon}_{j,z})$. Define the sign agreement frequency

$$q(J,Z) \equiv \frac{1}{JZ}\sum_{j,z} \mathbf{1}\left(\operatorname{sign}(Y^+)_{z,j} = \operatorname{sign}(\tilde{Y}^+)_{z,j}\right),$$

where $\mathbf{1}(\cdot)$ denotes the indicator function, and its population limit

$$q(J) \equiv \operatorname*{plim}_{Z\to\infty} q(J,Z).$$

Theorem 2 characterizes the limits of these population objects as $J \to \infty$. We also use simulations to validate our results away from these limits.

Theorem 2 (Sign Instability of Y^+) *Under Assumptions 3–5, for almost every realization of $(\alpha_j, \beta_j)_j$, the following hold.*

(i) ***Individual coin flip.*** *For each fixed asset j and state z,*

$$\lim_{J\to\infty} \pi(j,z) = \lim_{J\to\infty} P\left((\Sigma^{-1}y_z)_j > 0\right) = \frac{1}{2}.$$

Moreover, the distribution of $(\Sigma^{-1}y_z)_j$ is continuous and centered at zero in the limit, so the positive and negative tails are mirror images of equal magnitude.

(ii) ***Factor structure knowledge is insufficient.*** *Let Y and $\tilde{Y}$ be two payoff matrices generated by the same factor loadings $(\alpha_j, \beta_j)_j$ and factor realizations $(f_z)_z$ but independent idiosyncratic shocks $(\varepsilon_{j,z})$ and $(\tilde{\varepsilon}_{j,z})$. The population sign-determining vari-*

ables $(\Sigma^{-1}y_z)_j$ *and* $(\Sigma^{-1}\tilde{y}_z)_j$ *are asymptotically independent for each fixed* (z, j)*, each with limiting sign probability* $\frac{1}{2}$*. Consequently,*

$$\lim_{J\to\infty} q(J) = \frac{1}{2}.$$

Knowledge of statistical properties of the return process is thus *not* sufficient to guarantee a well-behaved mapping from preferences to asset holdings. To the contrary, the latent mapping is generally ill-conditioned, with the sign of any given element being a coin flip. Hence the misalignment between supply shocks and the ideal experiment is pervasive for realistic payoff processes, and cannot be corrected for by controlling for factor exposures.

Calibration and Numerical Exploration Our theoretical results consider the limit $J, Z \to \infty$. We now study the behavior outside the limit using Monte Carlo simulations with payoff parameters that generate a share of idiosyncratic risk roughly consistent with the empirical data. Concretely, we assume:[5]

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

Figure 2 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. The left panel depicts the proportion of times any individual element of the matrix Y^+ is positive.[6] Thus, given $Y > 0$, almost half the elements of Y^+ have the wrong sign. The right panel compares the signs of $(Y^+)_{z,j}$ and $(\tilde{Y}^+)_{z,j}$ where both Y and $\tilde{Y}$ are generated from the same factor model and are thus indistinguishable in practice. We again observe that the signs coincide only 50% of the time, implying that there is no systematic way to correct for these sign errors.

[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]For each $Z \in \{150, 300\}$, we vary the number of assets $J \in \{2, 4, \ldots, 100\}$. We took the average of 1000 runs (of the Monte Carlo simulations). Figure 2 also depicts the 95% confidence interval.

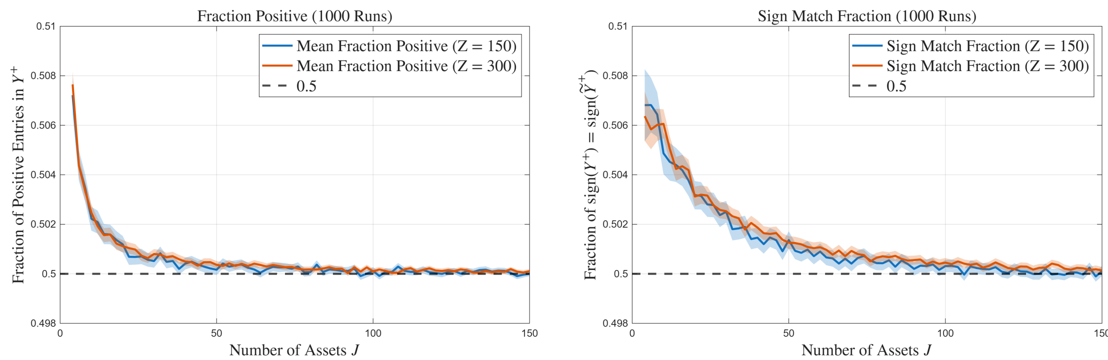


Figure 2: The Monte Carlo Simulation Results for Theorem 2. The left panel displays the empirical frequency of positive entries in Y^+, while the right panel shows the empirical frequency of sign matches between Y^+ and $\tilde{Y}^+$. Both panels report the results for $Z \in \{150, 300\}$ across various values of J, based on 1000 runs.

5 The Trilemma and its implications

We have established that the two principles of no arbitrage and preferences over payoffs sharply curtail the scope for non-parametric demand analysis in asset markets. We now summarize this result as a trilemma.

Theorem 3 (Trilemma) *Given observational data on portfolios, asset prices, and payoffs, one cannot jointly maintain (i) no-arbitrage asset pricing, (ii) investor preferences over payoffs, and (iii) model-free identification of structural asset demand functions.*

None of the stated conditions is easily discarded. No arbitrage is the prototypical internally-consistent pricing system, which ensures existence, consistency, and external validity of demand functions. Payoff-based asset valuation is the basic guiding principle of asset pricing. Since our definition of an asset is entirely generic, our results also apply equally to *portfolios* of primitive assets, which are themselves simply collections of payoffs. This leaves the assumption of constant payoffs. Unfortunately, this assumption is in tension with one of the basic functions of financial markets, which is price discovery—and thus revisions in expected payoffs—in response to news. It also cannot be directly verified in the data.

The trilemma is robust to small departures from its stated conditions. Relaxing no-arbitrage—for instance, by allowing for transaction costs or small arbitrage bands—replaces exact state-price equalities with inequalities but leaves the latent mapping Y^+ unobserved and ill-conditioned. Similarly, introducing non-pecuniary tastes or asset-level utility alters the fundamental demand object $\mathcal{D}^i$ but is orthogonal to the identification problem posed by Y^+: as long as investors retain any preference over state-contingent payoffs, the decomposition $A^i = (Y^+)^{\mathrm{T}}\mathcal{D}^i Y^+$ remains operative and Y^+ remains latent. The cross-asset linkages and sign instability documented in Theorem 2 depend only on the geometry of the payoff matrix, not on the preference specification. A model in which non-pecuniary considerations entirely dominate pecuniary ones is not, in any meaningful sense, a model of asset demand. Hence neither perturbation restores model-free identification of structural asset demand.

Our analysis suggests a critical role for structural models in asset demand estimation. However, since Y^+ cannot be identified from data, structurally estimated demand functions will depend sensitively on the assumed payoff structure. For example, Fuchs, Fukuda, and Neuhann (2025) show that the logit asset demand model proposed by Koijen and Yogo (2019) can yield low estimated demand elasticities because of the substitution patterns it assumes. The estimated demand elasticities should therefore be evaluated based on the validity and plausibility of the assumed payoff structure, not on empirical fit, which provides no information about whether the assumed latent mapping is correct. Our decomposition provides a framework for understanding what those restrictions imply and where misspecification is likely to be most consequential.

References

ALLEN, J., J. KASTL, AND M. WITTWER (2025): "Estimating Demand Systems with Bidding Data," Working Paper. [Cited on page 5.]

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

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 page 5.]

DUFFIE, D. (2001): Dynamic Asset Pricing Theory. Princeton University Press, third edn. [Cited on page 9.]

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 20.]

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 4.]

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 4.]

FUCHS, W., S. FUKUDA, AND D. NEUHANN (2025): "Demand-System Asset Pricing: Theoretical Foundations," Working Paper. [Cited on pages 3, 5, 22, 30, and 45.]

HADDAD, V., Z. HE, P. HUEBNER, P. KONDOR, AND E. LOUALICHE (2025): "Causal Inference for Asset Pricing," Working Paper. [Cited on pages 5 and 24.]

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 4.]

HE, Z., P. KONDOR, AND J. S. LI (2025): "Demand Elasticity in Dynamic Asset Pricing," Working Paper. [Cited on page 5.]

HURWICZ, L. (1962): "On the Structural Form of Interdependent Systems," in Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress, ed. by E. Nagel, P. Suppes, and A. Tarski, pp. 232–239. Stanford University Press. [Cited on page 16.]

IARIA, A., AND A. WANG (2020): "Identification and Estimation of Demand for Bundles," Working Paper. [Cited on page 4.]

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, 5, 9, and 30.]

KREPS, D. M., AND E. L. PORTEUS (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," Econometrica, 46(1), 185–200. [Cited on page 20.]

MARSCHAK, J. (1953): "Economic Measurements for Policy and Prediction," in Studies in Econometric Method, ed. by W. C. Hood, and T. C. Koopmans, no. 14 in Cowles Commission Monograph, pp. 1–26. John Wiley & Sons. [Cited on page 16.]

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.]

SHLEIFER, A. (1986): "Do Demand Curves for Stocks Slope Down?," Journal of Finance, 41(3), 579–590. [Cited on page 4.]

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 4.]

A Proofs of Propositions

A.1 Section 3

Proof of Proposition 1. First, the consumption process induced by portfolio a^i and payoff matrix Y is $c^i = Y^{\mathrm{T}} a^i + w^i$. Multiplying Y from the left, $Yc^i = YY^{\mathrm{T}} a^i + Yw^i$. Since YY^{T} is a $J \times J$ invertible matrix by Assumption 1, we have $a^i = (Y^+)^{\mathrm{T}} c^i - (Y^+)^{\mathrm{T}} w^i$. Since w^i does not depend on p, differentiating this expression yields $\frac{\partial a^i}{\partial p^{\mathrm{T}}} = (Y^+)^{\mathrm{T}} \frac{\partial c^i}{\partial p^{\mathrm{T}}}$. By no arbitrage, $p = Yq$ and hence $q = Y^+ p$. Then, equation (3) follows from the chain rule.

Second, under the stated conditions, the first-order condition with respect to c^i_z is:

$$\delta^i u'(c^i_z) = \lambda \pi_z^{-1} q_z,$$

where λ is the Lagrange multiplier on the budget constraint, which depends on the state price vector, as $\lambda = (1 - \delta^i) u'(c_0)$ where c_0 is optimal consumption at time 0. Differentiating this first-order condition with respect to $q_{z'}$ yields:

$$\delta^i u''(c^i_z) \frac{\partial c^i_z}{\partial q_{z'}} = \pi_z^{-1} \left(\frac{\partial \lambda}{\partial q_{z'}} q_z + \lambda \mathbf{1}(z = z') \right),$$

that is,

$$\frac{\partial c^i_z}{\partial q_{z'}} = \pi_z^{-1} \frac{1}{\delta^i u''(c^i_z)} \left(\frac{\partial \lambda}{\partial q_{z'}} q_z + \lambda \mathbf{1}(z = z') \right).$$

Letting $\tilde{\mathcal{D}}^i_{z,z'} \equiv \frac{1}{\delta^i u''(c^i_z)} \left(\frac{\partial \lambda}{\partial q_{z'}} q_z + \lambda \mathbf{1}(z = z') \right)$ and $\Pi \equiv \mathrm{diag}(\pi_1, \ldots, \pi_Z)$, we have:

$$\mathcal{D}^i = \Pi^{-1} \tilde{\mathcal{D}}^i.$$

Substituting this equation into (3) yields equation (4). ∎

Proof of Proposition 2. Let $Y \in \mathbb{R}^{J \times Z}$ be a payoff matrix, and let $(z_1, \ldots, z_T)$ be any sample of realized states. Let $y \in \mathbb{R}^J$ be any vector with $y \notin \{y(1), \ldots, y(Z)\}$, where $y(z)$ is the z-th column of Y. Define $\tilde{Y} = [Y \mid y] \in \mathbb{R}^{J \times (Z+1)}$, where y corresponds to a state that does not realize in the sample.

With these in mind, first, we show that $\tilde{Y}$ is observationally equivalent to

Y: the realized return on every asset j in every period t is identical under Y and $\tilde{Y}$. To see this, since the additional state corresponding to y does not realize in the sample, the realized return on asset j in period t is $(Y)_{j,z_t}$ under both Y and $\tilde{Y}$ for all $j \in \{1, \ldots, J\}$ and $t \in \{1, \ldots, T\}$.

Second, the Moore-Pensrose inverses of the two matrices differ: $\tilde{Y}^+ \neq Y^+$. On the one hand, the Moore-Penrose inverse $Y^+ \in \mathbb{R}^{Z \times J}$ is given by

$$Y^+ = Y^{\mathrm{T}}(YY^{\mathrm{T}})^{-1}.$$

On the other hand, the Moore-Penrose inverse $\tilde{Y}^+ \in \mathbb{R}^{(Z+1) \times J}$ is given by

$$\tilde{Y}^+ = \begin{bmatrix} Y^{\mathrm{T}}(YY^{\mathrm{T}} + yy^{\mathrm{T}})^{-1} \\ y^{\mathrm{T}}(YY^{\mathrm{T}} + yy^{\mathrm{T}})^{-1} \end{bmatrix}.$$

Thus, in addition to $Y^+ \neq \tilde{Y}^+$, the $Z \times J$ block of $\tilde{Y}^+$ is different from Y^+ as $(YY^{\mathrm{T}} + yy^{\mathrm{T}})^{-1} \neq (YY^{\mathrm{T}})^{-1}$. ■

Proof of Proposition 3. First, recall that $\mathcal{D}^i$ is the matrix of partial derivatives of optimal Marshallian consumption demand with respect to state prices, evaluated at given preferences and endowments. It is determined solely by the investor's preference ordering and budget set, and does not depend on Y directly. Hence, for any intervention that holds preferences and endowments fixed, $\mathcal{D}^i$ is unchanged.

In contrast, by Proposition 1, $\mathcal{A}^i = (Y^+)^{\mathrm{T}}\mathcal{D}^i Y^+$. Let $\tilde{Y}$ be a perturbation of Y, i.e., $\tilde{Y}$ is a $J \times \tilde{Z}$ non-negative matrix with $\mathrm{rank}(Y) = J \leq \tilde{Z}$. Let $\tilde{\mathcal{A}}^i = (\tilde{Y}^+)^{\mathrm{T}}\mathcal{D}^i\tilde{Y}^+$. Then, $\tilde{\mathcal{A}}^i \neq \mathcal{A}^i$ with probability 1 (also, the set of $\tilde{Y}$ with $\tilde{\mathcal{A}}^i \neq \mathcal{A}^i$ is open and dense).

As Proposition 2 establishes that Y is not identified from realized return data, $\mathcal{A}^i$ cannot be identified as a structural object, and that $\mathcal{A}^i$ alone does not suffice to predict asset demand responses to interventions without additional identifying assumptions on Y^+. This proves the first part, and the second part is a contraposition of the first part. ■

A.2 Section 4.1

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

$$Y^{+} = DY^{\mathrm{T}}. \tag{6}$$

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 $z \in \{1, \ldots, Z\}$, it follows that

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

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

$$0 = \sum_{z=1}^{Z} y_j(z) y_k(z). \tag{7}$$

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 (7) implies that, for any $z \in \{1, \ldots, Z\}$, there exists at most one $j \in \{1, \ldots, J\}$ 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 1 (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 $z \in \{1, \ldots, Z\}$, there exists at least one $j \in \{1, \ldots, J\}$ such that $Y_{j,z} = y_j(z) > 0$, it can be shown that the diagonal matrix D in expression (6) is unique.

Proof of Proposition 4. 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:

$$\sum_{z=1}^{Z} y_k(z) Y_{z,k}^{+} = 1 \quad \text{for all} \quad k \in \{1, \ldots, J\}; \tag{8}$$

$$\sum_{z=1}^{Z} y_j(z) Y_{z,k}^{+} = 0 \quad \text{if} \quad j \neq k. \tag{9}$$

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 (9) 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 (8). For the sum to be strictly positive, there must exist at least one index z^* such that:

$$y_k(z^*) > 0 \quad \text{and} \quad Y_{z^*,k}^{+} > 0. \tag{10}$$

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 (10), 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.3 Section 4.2

Proof of Proposition 5. Suppose for contradiction that $\mathcal{A}^i$ is identified but Y^+ is not assumed fixed. Then, there exist two observationally equivalent parameter vectors (Θ^i, Y, π) and $(\tilde{\Theta}^i, \tilde{Y}, \tilde{\pi})$ with $Y^+ \neq \tilde{Y}^+$. By Proposition 2, such pairs exist for any finite dataset. By the demand decomposition of Proposition 1,

$$\mathcal{A}^i(\Theta^i, Y, \pi) = (Y^+)^T \mathcal{D}^i Y^+ \neq (\tilde{Y}^+)^T \mathcal{D}^i \tilde{Y}^+ = \mathcal{A}^i(\tilde{\Theta}^i, \tilde{Y}, \tilde{\pi})$$

generically, contradicting observational equivalence. Hence $\mathcal{M}$ must include the assumption that Y^+ is fixed. The requirement of $K = J$ independent experiments then follows from the fact that $\mathcal{A}^i$ is a $J \times J$ matrix, and point-identification of all its elements requires J linearly independent price vectors in O_P. ■

A.4 Section 4.3: Proof of Theorem 2

The proof proceeds in 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 $\frac{1}{Z}Y^{\mathrm{T}}\Sigma^{-1}$. Then, the population limits $\pi(j,z)$ and $q(J)$ equal expressions involving the population inverse Σ^{-1}, reducing both parts of the theorem to statements about $(\Sigma^{-1}y_z)_j$. 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., $(\nu_{z,j})_{z,j}$ in the main text). Then, we show the sense in which the deterministic shift is small compared to the stochastic component $(\nu_{z,j})_{z,j}$ when J is large by applying the Woodbury identity to Σ. With these in mind, the third step establishes part(i). The proof shows that the sign of the (z,j) element of Y^+, which is approximated by that of $\Sigma^{-1}Y^T$ (up to scaling) is asymptotically a fair coin flip. Finally, the fourth step establishes part (ii). The proof shows that the signs for two economies with a share factor structure but independent realizations of idiosyncratic shocks are asymptotically independent.

Hereafter, fix a realization of (α, β) for which all laws of large numbers used below hold. Probabilities are conditional on the loadings unless noted otherwise.

Step 1. In the first step, we replace the sample Gram matrix $G_Z \equiv \frac{1}{Z} Y Y^{\mathrm{T}}$ 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, where, as in the main text,

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

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 2 below 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, we decompose $(\Sigma^{-1} y_z)_j$ into the deterministic shift and the stochastic part symmetric around 0. Writing $y_z = \gamma + \beta(f_z - \bar{f}) + \varepsilon_z$ as in the main text, one can express

$$(\Sigma^{-1} y_z)_j = \underbrace{(\Sigma^{-1}\gamma)_j}_{\mu_j} + \underbrace{(f_z - \overline{f})(\Sigma^{-1}\beta)_j + (\Sigma^{-1}\varepsilon_z)_j}_{\nu_{z,j}}. \tag{11}$$

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

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

$$\begin{aligned} P((\Sigma^{-1} y_z)_j > 0) &= P(\mu_j + \nu_{z,j} > 0) \\ &= 1 - F_{\nu,j}(-\mu_j) = \frac{1}{2} + f_{\nu,j}(0)\mu_j + O(\mu_j^2), \end{aligned} \tag{12}$$

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

With these in mind, we now establish Lemma 2, 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 2 (Population replacement) *Fix J and (α, β). As $Z \to \infty$, the sample Gram matrix $G_Z \equiv \frac{1}{Z}YY^{\mathrm{T}}$ converges almost surely to Σ and hence $G_Z^{-1} \to \Sigma^{-1}$ a.s. Moreover,*

$$\lim_{Z\to\infty} \max_{1\le j\le J} \left| \frac{1}{Z}\sum_{z=1}^{Z} \mathbf{1}\left(\left(\tfrac{1}{Z} y_z^{\mathrm{T}} G_Z^{-1}\right)_j > 0 \right) - P\left((\Sigma^{-1}y_z)_j > 0 \right) \right| = 0 \quad \textit{a.s.} \tag{13}$$

Consequently, conditional on (α, β),

$$\pi(j,z) = P\left((\Sigma^{-1}y_z)_j > 0 \right), \tag{14}$$

$$q(J) = \frac{1}{J}\sum_{j=1}^{J} P\left(\operatorname{sign}\big((\Sigma^{-1}y_z)_j\big) = \operatorname{sign}\big((\Sigma^{-1}\tilde{y}_z)_j\big) \right), \tag{15}$$

where y_z and $\tilde{y}_z$ denote the z-th columns of Y and $\tilde{Y}$.

Proof of Lemma 2. 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, $\|G_Z^{-1} - \Sigma^{-1}\|_F \le \eta$ a.s., so $|D_{z,j}| \le \frac{\eta}{Z}\|y_z\|$. A sign can flip only if $|(\frac{1}{Z}\Sigma^{-1}y_z^{\mathrm{T}})_j| \le |D_{z,j}|$. Since $(\Sigma^{-1}y_z)_j = \mu_j + \nu_{z,j}$ has a continuous density at around 0 with value $f_{\nu,j}(0)$, we have:

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

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 (13). Then, the representations (14)–(15) follow immediately. ■

Step 2. The second step examines the structure of Σ^{-1}. The key insight is that $\Sigma = \sigma_\varepsilon^2 I_J + UU^\mathrm{T}$ is a rank-two perturbation of a scaled identity. Applying the Woodbury identity yields $\Sigma^{-1} = \sigma_\varepsilon^{-2} I_J - \sigma_\varepsilon^{-4} U A_J^{-1} U^\mathrm{T}$, where $A_J \equiv I_2 + \sigma_\varepsilon^{-2} U^\mathrm{T} U$ is a 2×2 matrix with $\|A_J^{-1}\| = O(J^{-1})$, because $U^\mathrm{T} U$ grows like J as more assets are added. The correction term $U A_J^{-1} U^\mathrm{T}$ is therefore small relative to the leading $\sigma_\varepsilon^{-2} I_J$ term. This has the following consequences established in the lemma:

Lemma 3 (Small deterministic shift) *Under Assumptions 3–5, the following hold uniformly in j as $J \to \infty$:*

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

Consequently, the stochastic fluctuation $\nu_{z,j} \equiv (f_z - \bar{f})(\Sigma^{-1}\beta)_j + (\Sigma^{-1}\varepsilon_z)_j$ is symmetric around zero with a continuous density and variance satisfying $\sigma_{\nu,j}^2 \to \sigma_\varepsilon^{-2}$ uniformly in j.

Proof. We apply the Woodbury identity to $\Sigma = \sigma_\varepsilon^2 I_J + UU^\mathrm{T}$:

$$\Sigma^{-1} = \sigma_\varepsilon^{-2} \left(I_J - U \left(I_2 + \sigma_\varepsilon^{-2} U^\mathrm{T} U \right)^{-1} \sigma_\varepsilon^{-2} U^\mathrm{T} \right). \tag{16}$$

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 $j \in \{1, \ldots, J\}$, 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 $\left(I_2 + \sigma_\varepsilon^{-2} U^\mathrm{T} U\right)^{-1}$ is of order $O(J^{-1}) = O(1) \cdot O(J^{-1}) \cdot O(1)$. Consequently, each entry in the matrix $U\left(I_2 + \sigma_\varepsilon^{-2} U^\mathrm{T} U\right)^{-1} \sigma_\varepsilon^{-2} U^\mathrm{T}$ 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.

Since we have established $(I_2 + \sigma_\varepsilon^{-2} U^{\mathrm{T}} U)^{-1} = O(J^{-1})$, substituting this back into Woodbury identity (16) and noting that $\gamma = U v_1$ 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 $U^{\mathrm{T}}\gamma = U^{\mathrm{T}} U v_1$, which is $O(J)$. Thus, the second term equals

$$\sigma_\varepsilon^{-4} U \big(I_2 + \sigma_\varepsilon^{-2} U^{\mathrm{T}} U\big)^{-1} U^{\mathrm{T}} U v_1 = \sigma_\varepsilon^{-2} U \cdot O(J^{-1}) \cdot O(J) = \sigma_\varepsilon^{-2} U \cdot O(1).$$

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 (16) 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

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

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

Lemma 3 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 the density at zero governing the linearization of the sign probability is determined primarily by the idiosyncratic variance σ_ε^2.

Step 3: Individual coin flip, proof of part (i). By Lemma 2, it suffices to analyze $P\left((\Sigma^{-1}y_z)_j > 0\right)$. By (12), we have:

$$\pi(j,z) = \frac{1}{2} + f_{v,j}(0)\mu_j + O(\mu_j^2). \tag{17}$$

By Lemma 3, $\mu_j \to 0$ uniformly as $J \to \infty$. Lemma 3 also implies that the density $f_{v,j}(0)$ bounded uniformly in j as $J \to \infty$. Thus, $\lim_{J\to\infty} \pi(j,z) = \frac{1}{2}$ for each fixed (j,z).[7] This establishes part (i).

Step 4: Sign instability, part (ii). By Lemma 2, $q(J)$ equals the average over j of

$$P\left(\text{sign}((\Sigma^{-1}y_z)_j) = \text{sign}((\Sigma^{-1}\tilde{y}_z)_j)\right).$$

Since Y and $\tilde{Y}$ share the same factor structure, both Gram matrices converge to the same Σ. Applying decomposition (11) to each matrix at a fixed (z,j) gives

$$(\Sigma^{-1}y_z)_j = \mu_j + (f_z - \bar{f})(\Sigma^{-1}\beta)_j + (\Sigma^{-1}\varepsilon_z)_j;$$
$$(\Sigma^{-1}\tilde{y}_z)_j = \mu_j + (f_z - \bar{f})(\Sigma^{-1}\beta)_j + (\Sigma^{-1}\tilde{\varepsilon}_z)_j.$$

Both variables share the deterministic shift μ_j and the common factor component $(f_z - \bar{f})(\Sigma^{-1}\beta)_j$. By Lemma 3, $(\Sigma^{-1}\beta)_j = O(J^{-1})$, and thus the common factor component vanishes as $J \to \infty$. The residual stochastic terms $(\Sigma^{-1}\varepsilon_z)_j$ and $(\Sigma^{-1}\tilde{\varepsilon}_z)_j$ are independent (since $\varepsilon \perp \tilde{\varepsilon}$) and each symmetric around zero. In the limit, the two sign-determining variables reduce to independent symmetric random variables ξ and $\tilde{\xi}$. Hence,

$$P\left(\text{sign}((\Sigma^{-1}y_z)_j) = \text{sign}((\Sigma^{-1}\tilde{y}_z)_j)\right) \to P(\xi > 0, \tilde{\xi} > 0) + P(\xi < 0, \tilde{\xi} < 0)$$
$$= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

[7] An earlier working paper shows that $\frac{1}{J}\sum_{j=1}^{J} f_{v,j}(0)\mu_j \to \frac{1}{J} f_v(0)\Theta_1$, where the constant Θ_1 is determined by the model primitives.

Since this holds for each fixed j and the indicators are bounded, the dominated convergence theorem gives $\lim_{J\to\infty} q(J) = \frac{1}{2}$. ■

To conclude the proof of Theorem 2, two remarks are in order. Remark 2 shows that a multi-factor extension is possible. Remark 3 reiterates the economic interpretation of Theorem 2.

Remark 2 (Multi-factor extension) *For a K-factor model with*

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

the population Gram matrix takes the form

$$\Sigma = \sigma_\varepsilon^2 I_J + UU^{\mathrm{T}},$$

where

$$U \equiv \begin{bmatrix} \gamma & \sigma_{f,1}\beta^{(1)} & \dots & \sigma_{f,K}\beta^{(K)} \end{bmatrix} \in \mathbb{R}^{J\times(K+1)}.$$

Since $(I_{K+1} + \sigma_\varepsilon^{-2} U^{\mathrm{T}} U)^{-1} = O(J^{-1})$ for any fixed K, Lemma 3 carries over verbatim: both $(\Sigma^{-1}\gamma)_j$ and each $(\Sigma^{-1}\beta^{(k)})_j$ are $O(J^{-1})$. Hence both parts of Theorem 2 hold unchanged for any finite number of factors.

Remark 3 (Economic interpretation) *Theorem 2 sharpens the identification impossibility established in Theorem 1 of the paper. Theorem 1 shows that supply shocks generically fail to generate the correct direction of state-price changes for at least one asset. Part (i) here quantifies how severe this directional failure is for each state-asset pair individually: the sign of the required correction is a coin flip in large economies. Part (ii) shows that this instability cannot be resolved by obtaining a second sample from an economy with the same factor structure: even sharing the same systematic risk, two economies require corrections of opposite sign approximately half the time. This rules out any procedure for controlling directional errors that relies solely on the factor structure of payoffs.*

B Illustration in a General Equilibrium Model

We illustrate our findings using a simple example economy based on Fuchs, Fukuda, and Neuhann (2025). The decision problem is as in (PCP). For tractability, we assume that all investors are symmetric, face no portfolio constraints and have log utility.

Payoffs. There are two assets and two aggregate states of the world, both denoted by g (green) and r (red). The probability of state $z \in \{g, r\}$ is $\pi_z \in (0,1)$. Table 1 depicts the payoff matrix. Parameter $\epsilon \in (0,1)$ determines the complementarity between green and red assets. As $\epsilon \to 0$, green and red assets are perfect substitutes. As $\epsilon \to 1$, 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 1: Payoff matrix.

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 which we use to create price variation.

Asset demand. We will be interested in analyzing asset demand functions in a neighborhood around $s_g = 0$. We derive the demand functions a_g and a_r directly from the representative agent's portfolio choice problem:

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

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

$$(1-\delta)\frac{p_g}{E_0 - p_g(a_g - E_g) - p_r(a_r - E_r)} = \delta\pi_g \frac{1+\epsilon}{(1+\epsilon)a_g + (1-\epsilon)a_r} + \delta\pi_r \frac{1-\epsilon}{(1-\epsilon)a_g + (1+\epsilon)a_r}; \tag{18}$$

$$(1-\delta)\frac{p_r}{E_0 - p_g(a_g - E_g) - p_r(a_r - E_r)} = \delta\pi_g \frac{1-\epsilon}{(1+\epsilon)a_g + (1-\epsilon)a_r} + \delta\pi_r \frac{1+\epsilon}{(1-\epsilon)a_g + (1+\epsilon)a_r}. \tag{19}$$

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

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

$$a_g(p_g, p_r) = \delta\frac{(1 + p_g(1+s_g) + p_r)\left((1-\epsilon^2)p_g - ((1+\epsilon)^2 - 4\epsilon\rho)p_r\right)}{(p_g - p_r)^2 - (p_g + p_r)^2\epsilon^2}. \tag{20}$$

Varying only the green assets' price yields the standard own-price elasticity:

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

Misalignment between ideal experiment and supply shock. In the ideal experiment, the investor faces an exogenous increase in the price of the green asset p_g while p_r remains fixed. By Lemma 1, the induced change in state prices is

$$\Delta\mathbf{q}_g^{\text{ideal}} = \frac{\partial}{\partial p(g)}\begin{bmatrix} q(g) \\ q(r) \end{bmatrix} = \frac{1}{2\epsilon}\begin{bmatrix} 1+\epsilon \\ -(1-\epsilon) \end{bmatrix}. \tag{21}$$

A pure shock to $p(g)r$ thus *raises* the state price in state g, but *lowers* it in state r. This decrease in $q(r)$ is necessary to keep $p(r)$ unchanged.

Now consider how a supply shock s_g affects equilibrium state prices. Given resource constraints $c(z) = y_g(z)(1+s_g) + y_r(z)$, equilibrium state prices are

$$q(g) = \pi_g \frac{\delta}{1-\delta} \cdot \frac{1}{1+\frac{1+\epsilon}{2}s_g} \qquad \text{and} \qquad q(r) = \pi_r \frac{\delta}{1-\delta} \cdot \frac{1}{1+\frac{1-\epsilon}{2}s_g}. \tag{22}$$

Differentiating yields

$$\Delta \mathbf{q}_g^{\text{supply}} = \frac{\partial}{\partial s_g}\begin{bmatrix} q(g) \\ q(r) \end{bmatrix} = -\frac{1-\delta}{2\delta}\begin{bmatrix} (1+\epsilon)\cdot \frac{q(g)^2}{\pi_g} \\ (1-\epsilon)\cdot \frac{q(r)^2}{\pi_r} \end{bmatrix} < 0. \tag{23}$$

In contrast to the ideal experiment, a positive supply shock to the green asset thus decreases *both* state prices whenever $\epsilon < 1$. The simple reason is that the green asset pays off in both states of the world. As such, the supply shock generates a state price change Δq_g that is of the *wrong sign* compared to the ideal experiment. The only exception is when both assets are Arrow securities ($\epsilon = 1$).

Implications for demand elasticties. The misalignment between supply shock and ideal experiment can sharply bias observed behavior. In the ideal experiment, the investor is more willing to substitute away from green assets because the price of the red asset is unchanged. In the supply shock, substitution is tempered because the red asset is endogenously repriced. The resulting "elasticity" measure $\mathcal{E}_g^{\text{supply}}$ thus has an additional term which accounts the spillover to p_r:

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

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

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

We plot both measures in Figure 3. The two differ by order of magnitude for small ϵ. In this range, the two assets are close substitutes. In the ideal experiment

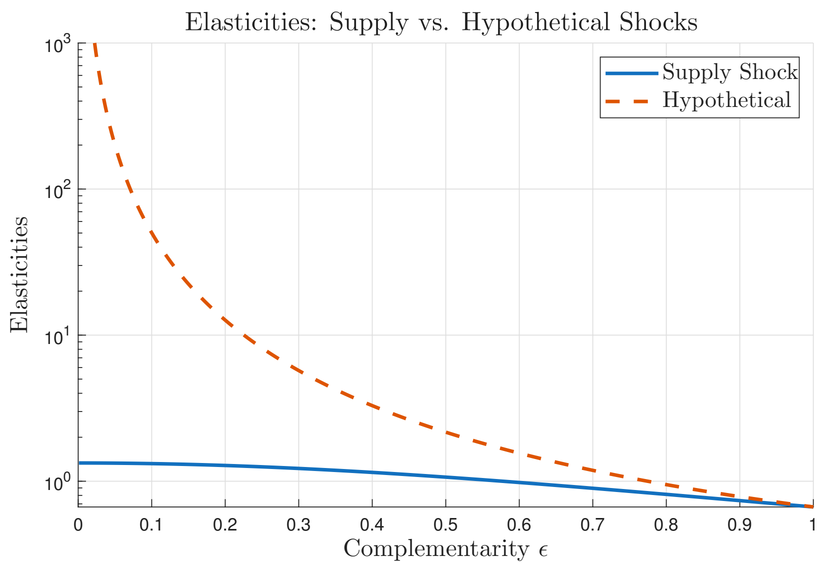


Figure 3: Ideal vs. supply-shock elasticities as a function of ϵ for $\delta = 2/3$ and $\pi_r = \pi_g = 1/2$. The ideal elasticity (solid line) diverges as $\epsilon \to 0$, while the supply-shock elasticity (dashed line) remains bounded. Both elasticities converge to $1 - \delta\pi_g = 2/3$ at the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

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. The only exception is when $\epsilon \to 1$ and the assets approach Arrow securities. In this case, there is no spillover across assets and thus no difference between the ideal experiment and the supply shock.

C Empirical Illustration

To further gauge the empirical relevance of our results, we conduct a simple empirical exercise in which we use payoff data from the *S&P* 500 to assess the alignment between (subsets of) the payoff matrix and its inverse. The exercise is not intended to be exhaustive, but simply illustrates the immediate relevance of the issues we discuss. The sample consists of 428 stocks that remained in the *S&P* 500 from 2020 to 2024. Since the true payoff matrix is latent, we construct (subsets) of it by sampling realized payoffs.

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 repeat this exercise ten times with replacement and report averages across all ten repetitions. Table 2 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. Taken together, 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 2: Results of our empirical exercise averaged over 10 iterations.

D Online Appendix

The Online Appendix is structured as follows. Appendix D.1 presents an example where redundant assets cause discontinuous demand. Appendix D.2 supplements Section 3 (specifically, Proposition 3). Appendix D.3 complements Section 4.1 by providing conditions under which Y^+ has the wrong sign for each state (Proposition 6), analogous to the asset-specific conditions in Proposition 4.

D.1 Section 2.2

We present an example in which an asset demand function exhibits discontinuity in the presence of redundant assets.

Example 3 (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.

1. *Suppose $\Phi^i = \mathbb{R}^3$. 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 (PCP) is no longer well-defined, and well-defined demand functions no longer exist.*

2. *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.

D.2 Section 3

Remark 4 (Proposition 3) *Supposing that $\mathcal{D}^i$ is invertible, if $\text{row}(Y) \neq \text{row}(\tilde{Y})$ then $\mathcal{A}^i \neq \tilde{\mathcal{A}}^i$.*

To see this, for ease of notation, we introduce matrices $M_Y \equiv (Y^+)^T \mathcal{D}^i Y^+$ and $M_{\tilde{Y}} \equiv (\tilde{Y}^+)^T \mathcal{D}^i \tilde{Y}^+$. We have $\mathcal{A}^i \neq \tilde{\mathcal{A}}^i$ if (and only if) $M_Y \neq M_{\tilde{Y}}$.

Since $(YY^\mathrm{T})^{-1}$ is an invertible $J \times J$ matrix, the range of $(Y^+)^\mathrm{T}$ satisfies:

$$\text{Range}((Y^+)^\mathrm{T}) = \text{Range}(Y^\mathrm{T}(YY^\mathrm{T})^{-1}) = \text{Range}(Y^\mathrm{T}) = \text{row}(Y).$$

Then, we consider the full product M_Y. Since $\mathcal{D}^i$ is assumed to be invertible and Y^+ has full rank J, the product $\mathcal{D}^i Y^+$ is a $J \times Z$ matrix with rank J. Thus, invoking the properties of the range of a matrix product,

$$\text{Range}(M_Y) = \text{Range}((Y^+)^\mathrm{T}) = \text{row}(Y).$$

Now, suppose toward a contradiction that $M_Y = M_{\tilde{Y}}$. If two matrices are equal, then: $\text{Range}(M_Y) = \text{Range}(M_{\tilde{Y}})$. Substituting our previous result:

$$\text{row}(Y) = \text{row}(\tilde{Y}).$$

This contradicts the initial hypothesis that $\text{row}(Y) \neq \text{row}(\tilde{Y})$. Hence, $M_Y \neq M_{\tilde{Y}}$, and consequently $\mathcal{A}^i \neq \tilde{\mathcal{A}}^i$. The proof also states if $\text{row}(Y^+) \neq \text{row}(\tilde{Y}^+)$ then $\mathcal{A}^i \neq \tilde{\mathcal{A}}^i$.

D.3 Section 4.1

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

Proposition 6 *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* $z \in \{1, \dots, Z\}$*, 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 6, 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 6. 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

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

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 $z \in \{1, \dots, Z\}$, 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

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

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). ∎