On the Recovery of Demand Elasticities in Dynamic Settings*

Carter Davis† Mahyar Kargar‡ Jiacui Li§ Dejanir Silva¶

August 27, 2026

(first draft: April 16, 2016)

Abstract

We study how to recover structural demand elasticities in dynamic asset markets. An identifying shock moves current prices together with expected returns, dividends, and risk, so the instrumented elasticity depends on the shock. We formulate a sequence-space demand system that nests static and dynamic models. In static economies, cross-asset price multipliers recover structural demand when price responses span the price space. In dynamic economies, our Recovery Theorem uses response paths across horizons to recover the demand kernel under a dynamic rank condition. We implement the approach with an orthonormal factor model. In weekly order-flow data, the naive elasticity is about 0.35, compared with a recovered own-price elasticity of about 5; systematic demand responds to expected returns throughout the following month.

Keywords: demand elasticity, price impact, demand system asset pricing

JEL Classification: G11, G12, G14.

*We thank Ehsan Azarmsa, Aditya Chaudhry, William Fuchs, Xavier Gabaix, Andrei Gonçalves, Valentin Haddad, Péter Kondor, Erik Loualiche, Tyler Muir, Daniel Neuhann, Zhaoxin Shi, Andrea Tamoni, Philippe van der Beck, and conference and seminar participants at NBER New Developments in Long-Term Asset Management Spring 2026 Meeting, SFS Cavalcade 2026, WFA 2026, 2026 Wabash River Finance Conference, Boston College, Carnegie Mellon University, and The Ohio State University, for their helpful comments and suggestions.

†Fisher College of Business, The Ohio State University; carter.davis@fisher.osu.edu

‡Gies College of Business, University of Illinois Urbana-Champaign; kargar@illinois.edu

§David Eccles School of Business, University of Utah; jiacui.li@eccles.utah.edu

¶Mitchell Daniels School of Business, Purdue University; dejanir@purdue.edu

1 Introduction

Investor demand has become central to understanding equilibrium asset prices and conducting counterfactual analysis. Demand system asset pricing (DSAP) turns this idea into an empirical framework by estimating demand systems and using them to map shifts in holdings into equilibrium prices. This approach has been applied to a wide range of topics, including ESG mandates (Koijen, Richmond, and Yogo, 2024; Van der Beck, 2021), passive intermediation (Haddad, Huebner, and Loualiche, 2025b), household participation (Davis, Knüpfer, Soerlie Kvaerner, Sen Dogan, and Vokata, 2024), and anomaly returns (Tamoni, Sokolinski, and Li, 2024). Yet the demand systems estimated in this literature are typically static: they describe how holdings respond to current prices and investor or asset characteristics. Portfolio theory generally implies a richer demand system in which current holdings also depend on expected future prices, cash flows, and risk. Bridging this gap requires a way to estimate dynamic demand systems that more closely reflect the portfolio-choice models underlying modern asset pricing.

This paper develops a method to estimate a dynamic asset-demand system with rich cross-asset substitution and forward-looking dependence on future prices, cash flows, and risk. The system is flexible enough to nest linearized versions of a broad class of canonical portfolio-choice models.[1] Applying the method to weekly U.S. stock order flow, we find that accounting for dynamics raises the estimated own-price elasticity from approximately 0.35 under the conventional static interpretation to approximately 5. This difference suggests that incorporating dynamics can substantially alter both our inference about investor demand and the resulting counterfactuals.

Performing counterfactual analysis often requires recovering not only the contemporaneous response to changes in prices, but the entire *structural demand kernel*: the collection of coefficients

[1]These include static demand systems (Koijen and Yogo, 2019), myopic mean–variance demand, intertemporal hedging demand (Merton, 1973; Campbell and Viceira, 2002), dynamic adjustment costs (Gârleanu and Pedersen, 2013), and models with belief distortions or dispersed information (Bordalo, Gennaioli, and Shleifer, 2018).

governing how current holdings respond separately to current prices, expected future prices and cash flows, and risk. Each coefficient answers a ceteris paribus question, defined by changing one demand argument while holding the others fixed. These coefficients are structural in the sense that they remain invariant when interventions change the process governing an outside investor's inelastic holdings but leave other investors' preferences and constraints unchanged. This invariance allows demand estimates to be transported across shock environments.

Estimating this richer demand system, however, presents three econometric challenges. First, instruments that induce exogenous price variation do not generally isolate individual demand arguments. A shock to one asset can move other assets through substitution and no-arbitrage linkages (Fuchs, Fukuda, and Neuhann, 2023; Haddad, He, Huebner, Kondor, and Loualiche, 2025a), while a persistent shock moves current prices together with expected future prices and returns (Davis, Kargar, and Li, 2025; Binsbergen, David, and Opp, 2025). Expected returns and risk may also respond together, making an observed demand slope a mixture of distinct channels (He, Kondor, and Li, 2025). Second, investors may have access to trading signals or information unavailable to the econometrician, so the conditional expectations entering demand are latent. Third, dynamics add endogenous demand arguments and structural coefficients, but we do not give the econometrician additional external instruments beyond those needed to identify demand in the static setting.

These challenges imply that methods designed for static demand systems generally do not recover the structural coefficients of dynamic demand. Rather than require instruments that isolate individual demand arguments, we exploit the joint equilibrium responses they generate across assets and horizons. In the simplest static setting without cross-asset substitution, the structural price elasticity can be recovered from the inverse of a price multiplier, the equilibrium response of prices to an instrument-induced shift in outside demand. With dynamics, this equivalence survives

only when the identifying variation is transitory and leaves future demand arguments unchanged. A persistent shift moves not only the current price but also expected future prices, so the impact price multiplier combines the contemporaneous and forward-looking coefficients of the structural demand kernel.

Our approach recovers these coefficients using *intertemporal price multipliers*, the equilibrium price responses to the instrument across multiple horizons. Market clearing at each date generates a set of *dynamic recovery equations* linking these observable responses to the structural demand kernel. Stacking these equations across assets and horizons recovers the kernel when the resulting response paths satisfy a rank condition. Crucially, these additional restrictions do not require additional external instruments: the instruments available in the static setting generate an entire path of equilibrium responses, and the dynamic recovery equations use variation along this path to identify the additional coefficients. Thus, the same persistence that invalidates the static interpretation of the impact multiplier generates the additional equilibrium restrictions needed for dynamic recovery. In this sense, persistence is a blessing, not a curse.

Persistence alone, however, is not sufficient for identification. Under an AR(1) response, current and future prices move in fixed proportions, so the data identify only a combination of contemporaneous and forward-looking coefficients. Recovery requires response paths with enough independent variation across assets or horizons.

This recovery procedure relies on representing demand in *sequence space*. Dynamic portfolio models are often written in state space, with demand expressed as a function of state variables that summarize investors' forecasts. Estimating such a demand rule requires the econometrician to specify the relevant state vector, which may include signals unavailable to the econometrician. In sequence space, we instead write demand directly as a function of expected future paths. Iterated expectations allow us to recover the projections of these latent expectations from observable future

realizations, even when investors have access to richer information than the econometrician. The sequence-space representation also separates the demand kernel from the processes governing the state variables, preserving the kernel across changes in the outside-demand process that would generally alter the state-space representation. Together, the sequence-space representation and the recovery equations allow empirical researchers to incorporate forward-looking demand without requiring additional external instruments, while allowing theoretical researchers to take models with rich dynamics to the data without reducing them to static demand systems.

To take the population recovery result to stock-level data, we develop a two-step estimator. We first estimate the responses of the demand arguments to the instruments using local projections and then treat these response coefficients as generated data in a minimum-distance recovery step. A bootstrap propagates the joint estimation uncertainty from the local projections through the recovered demand coefficients. To accommodate a large, unbalanced stock panel, we impose an orthonormal factor representation that preserves an own-stock component while compressing systematic substitution and risk channels. The factor structure reduces the dimension of the problem and supplies the restrictions needed for identification, consistent with the limits of factor-based demand identification emphasized by Baba Yara, Boyer, and Davis (2021) and Fuchs, Fukuda, and Neuhann (2025).

The empirical results show that dynamic recovery is quantitatively important. In weekly U.S. stock order-flow data, the naive inverse-multiplier benchmark implies an own-price elasticity of about 0.35, while the recovered stock-specific own-price elasticity is about 5. In the preferred four-week specification, the stock-specific component is concentrated in the first week, but systematic factor demand responds to expected returns throughout the following month. Controlling for the shock-induced risk response raises the first-week elasticity by a precisely estimated amount, and the combined risk loading beyond week one is positive in 97% of bootstrap draws. Thus, the

same identified innovation reveals both a strong immediate own-stock response and economically important forward-looking systematic demand.

The rest of the paper is organized as follows. Section 2 develops the structural sequence-space demand system, nests canonical models, and states the econometric challenges. Section 3 recovers the structural demand matrix from cross-asset multipliers in a static economy. Section 4 proves the Recovery Theorem using intertemporal price multipliers and characterizes the dynamic-rank condition. Section 5 develops the two-step estimator and adapts it to a high-dimensional, unbalanced stock panel using an orthonormal factor representation. Section 6 presents the empirical implementation and results. Section 7 concludes.

2 The Dynamic Demand System

Modern portfolio theory suggests that investors' portfolios depend not only on current prices, but also on the expected paths of future prices, cash flows, and risk. Accordingly, empirical demand models should be sufficiently rich to capture this dependence. In this section, we formulate a *dynamic demand system* that nests several canonical models of portfolio demand. This added richness brings demand estimation closer to theory, but also creates additional challenges for identifying the system from observed data. We delineate the resulting identification problem, while Sections 3 and 4 develop approaches for addressing it.

2.1 Environment and dynamic demand

Assets. Consider a discrete-time economy with N risky assets indexed by $j = 1, \ldots, N$ and a risk-free asset with constant gross return R^f. Let $P_t \in \mathbb{R}^N$ and $\mathcal{D}_t \in \mathbb{R}^N$ denote the vectors of prices and dividends at time t, and $r_{t+1} \in \mathbb{R}^N$ denote the vector of one-period log returns.

Let $\nu_t = (\nu_{t,1}, \ldots, \nu_{t,K})$ denote a K-dimensional vector of shocks at time t, and let $\nu^t = (\nu_0, \nu_1, \ldots, \nu_t)$ denote the history of shocks up to time t. The vector ν_t collects all primitive sources of randomness in the economy, including signals observed by investors but potentially unavailable to the econometrician. We take dividends as exogenous, with $\mathcal{D}_t = \mathcal{D}(\nu^t, t)$.

Investors and demand. Let $Q_t \in \mathbb{R}^N$ denote the portfolio holdings of a representative elastic investor. Fix a reference point $(\bar{Q}, \bar{P}, \bar{\mathcal{D}})$ with strictly positive components and define log deviations[2]

$$q_t \equiv \log Q_t - \log \bar{Q}, \qquad p_t \equiv \log P_t - \log \bar{P}, \qquad d_t \equiv \log \mathcal{D}_t - \log \bar{\mathcal{D}}.$$

For each horizon $s \geq 1$, define the conditional first moments

$$\mu^p_{t,t+s} \equiv \mathbb{E}_t[p_{t+s}], \qquad \mu^d_{t,t+s} \equiv \mathbb{E}_t[d_{t+s}].$$

Here and throughout, $\mathbb{E}_t$ denotes rational conditional expectations given investors' information at date t. The econometrician need not observe that information set, the state variables that summarize it, or the expectations themselves. Define the return covariance matrix $V^r_{t,t+s} \equiv \mathbb{V}\mathrm{ar}_t(r_{t+s})$, which is a symmetric $N \times N$ matrix. Let $m \equiv N(N+1)/2$ and define its half-vectorization $v^r_{t,t+s} \equiv \mathrm{vech}(V^r_{t,t+s}) \in \mathbb{R}^m$. Working with the $\mathrm{vech}(\cdot)$ operator ensures that the second moments are vectors of m unique covariance terms rather than redundant $N \times N$ matrices.

We allow current demand to depend on the current price and on the expected paths of future prices, dividends, and conditional return risk. Throughout the paper, we work with the log-

[2]For simplicity, we assume that prices, dividends, and holdings are stationary, so there is a well-defined time-invariant reference point. It is straightforward to generalize the results to the case where only variables scaled by a stochastic trend, such as the price-earnings ratio, are stationary.

linearized demand system

$$q_t = -A_0\, p_t + \sum_{s=1}^{S} A_s\, \mu^p_{t,t+s} + \sum_{s=1}^{S} B_s\, \mu^d_{t,t+s} - \sum_{s=1}^{S} C_s\, \nu^r_{t,t+s} + \xi_t, \tag{1}$$

where $A_0, A_s \in \mathbb{R}^{N\times N}$, $B_s \in \mathbb{R}^{N\times N}$, and $C_s \in \mathbb{R}^{N\times m}$. The matrix A_0 governs the contemporaneous response of holdings to current prices, holding fixed the expected paths of future prices, dividends, and risk. We assume that A_0 is invertible. The matrices A_s, B_s, and C_s capture how expectations of future prices, dividends, and risk at horizon s influence current demand, holding the other arguments fixed. We refer to the collection

$$\mathcal{K} \equiv \left\{A_0, \{A_s, B_s, C_s\}_{s=1}^{S}\right\}$$

as the *dynamic demand kernel*. The contemporaneous price elasticity encoded in A_0 is an economically important component of $\mathcal{K}$, but counterfactuals generically depend on the entire kernel. The vector ξ_t is a latent demand shifter, unobserved and potentially correlated with equilibrium prices, in the sense of Koijen and Yogo (2019).[3] We allow the latent demand shifter to depend on the history of primitive shocks: $\xi_t = \xi(\nu^t, t)$.

Outside investor and market clearing. Let $Z_t \in \mathbb{R}^N$ denote the (inelastic) holdings of an outside investor, with reference point $\bar{Z}$. Outside holdings are an exogenous function of the shocks: $Z_t = Z(\nu^t, t)$. Total shares outstanding are fixed and normalized so that

$$Q_t + Z_t = \mathbf{1}_N.$$

[3] Appendix M shows how the latent demand ξ_t may capture belief shocks, noisy signals, shocks to preferences or financial constraints that are not directly observable by an econometrician.

Linearizing around $(\bar{Q},\bar{Z})$ with $\bar{Q} + \bar{Z} = \mathbf{1}_N$, we can write the market-clearing condition as

$$q_t = -z_t, \qquad z_t \equiv \text{Diag}(\bar{Q})^{-1}\left(Z_t - \bar{Z}\right). \tag{2}$$

We express outside-demand deviations in units of the elastic investors' reference holdings. Under this normalization, a one-unit increase in outside demand reduces elastic investors' holdings by one unit. Equivalently, $-z_t$ is the *residual-supply* deviation absorbed by elastic investors.

An equilibrium requires the demand system (1) and the market-clearing condition (2) to hold jointly.

Definition 1 (Equilibrium)**.** *Given exogenous processes for normalized outside demand $\{z_t\}$, latent demand $\{\xi_t\}$, and dividends $\{d_t\}$, an equilibrium is a pair of stochastic processes $\{p_t, q_t\}$ adapted to the filtration generated by v^t such that:*

1. Demand optimality. *The investor's holdings satisfy the demand system* (1) *at every date t.*
2. Market clearing. *The market clearing condition $q_t = -z_t$ holds at every date t.*

2.2 Structural demand in sequence space

The demand system (1) is the primitive object of our analysis.[4] This subsection explains the sense in which its coefficients are structural and why we represent the system in sequence space rather than state space.

A structural demand system. The demand system in equation (1) is structural: its coefficients describe portfolio responses under counterfactual changes in its arguments. Each coefficient

[4] Appendix M derives the demand system from a general portfolio problem, and Appendix G discusses the connection between demand and preferences.

answers a *ceteris paribus* question: it is defined by changing one argument while holding the remaining arguments fixed. For example, A_0 measures the response of demand to current prices holding investors' expectations about future variables fixed. The structural coefficients therefore differ from the equilibrium comovement between holdings and prices.

These demand coefficients are structural in the sense that they remain invariant across a particular class of interventions. An intervention may change other parts of the equilibrium system, and therefore the resulting equilibrium relationships, without changing the demand coefficients (Haavelmo, 1943). As emphasized by Hurwicz (1962), a coefficient is structural only relative to a specified class of interventions.[5] We consider interventions that change the stochastic process for outside demand while leaving investors' preferences and constraints unchanged. The dynamic demand kernel $\mathcal{K}$ is invariant across this class of interventions. Interventions that directly change preferences or constraints may, of course, change the kernel.

Alternative arguments of demand. The choice of arguments determines the counterfactual interpretation of the coefficients. For example, demand may instead be written in terms of expected returns:

$$q_t = -\sum_{s=1}^{S} C_s \, v^r_{t,t+s} + \sum_{s=1}^{S} D_s \, \mathbb{E}_t[r_{t+s}] + \xi_t. \tag{3}$$

In this representation, D_s measures the response of current holdings to the expected return at horizon s, holding the other arguments fixed. Using the Campbell–Shiller approximation

$$r_{t+1} = \kappa + \rho p_{t+1} + (1-\rho) d_{t+1} - p_t,$$

[5] See also Heckman and Pinto (2024) for a related discussion of autonomous structural equations.

the return-based system imposes the restrictions

$$A_0 = D_1, \qquad A_s = \rho\, D_s - D_{s+1} \ \text{ for } s = 1, \ldots, S, \qquad B_s = (1-\rho)\, D_s \ \text{ for } s = 1, \ldots, S,$$

where $D_{S+1} \equiv 0$ and constants are absorbed into ξ_t. Thus, alternative representations need not define different economic models, but their coefficients answer different counterfactual questions because they hold different objects fixed.

Sequence-space and state-space representations. Equation (1) represents current demand as a function of sequences of conditional moments. This is a *sequence-space representation* of the demand system.[6] Dynamic portfolio-choice models are instead commonly solved in state space, yielding a portfolio rule as a function of a state vector.[7]

The two representations organize the same investor problem differently, but they need not have the same invariance properties. The sequence-space kernel maps a given set of conditional moments into current demand. A state-space portfolio rule substitutes equilibrium laws of motion for those conditional moments and thereby combines the demand kernel with the processes governing the state variables. Changing the process for outside demand generally changes equilibrium prices and their conditional moments. It therefore changes the state-space portfolio rule even when the underlying sequence-space kernel remains fixed.

A simple example illustrates the distinction. Consider the scalar dynamic demand system

$$q_t = -a_0 p_t + a_1 \mathbb{E}_t[p_{t+1}] + \xi_t. \tag{4}$$

[6] See Ljungqvist and Sargent (2018) for a discussion of sequence-space and state-space representations of dynamic models, and Appendix H for a detailed discussion in our setting.

[7] The dynamic programming approach delivers portfolio rules as functions of state variables; see, for example, Merton (1969, 1973) and Campbell and Viceira (2002). The martingale approach of Karatzas, Lehoczky, and Shreve (1987) and Cox and Huang (1989) is closer in spirit to the sequence-space representation, but recovering the portfolio policy that implements the optimal wealth profile often requires specifying the state dynamics.

In this case, the dynamic demand kernel is characterized by the two structural coefficients (a_0,a_1).

Suppose that the initial outside-demand process implies prices follow an AR(1) process:

$$\mathbb{E}_t[p_{t+1}] = \rho p_t. \tag{5}$$

Substituting (5) into (4) gives the state-space portfolio rule

$$q_t = -(a_0 - \rho a_1)p_t + \xi_t.$$

Its coefficient combines the structural demand coefficients (a_0,a_1) with the equilibrium persistence parameter ρ into a single coefficient $(a_0 - \rho a_1)$ on the current price. Hence, the coefficient of the state-space representation depends not only on the structural demand coefficients, but also on the equilibrium price process, as summarized by the persistence parameter ρ. The equilibrium price process, in turn, depends on the stochastic process for outside demand.

Now consider an intervention that changes the outside-demand process and induces an AR(2) price process,

$$\mathbb{E}_t[p_{t+1}] = \rho_1 p_t + \rho_2 p_{t-1}.$$

The resulting state-space portfolio rule becomes

$$q_t = -(a_0 - \rho_1 a_1)p_t + \rho_2 a_1 p_{t-1} + \xi_t.$$

The intervention changes both the coefficients and the arguments of the state-space rule. By contrast, the sequence-space coefficients (a_0,a_1) remain unchanged. For this class of interventions, the sequence-space representation is therefore autonomous in the sense of Hurwicz (1962), whereas the state-space representation is not.

2.3 Nesting canonical models

The dynamic demand system introduced above is flexible enough to capture leading portfolio models, which differ in both the variables that shape expectations and the time horizons over which those expectations influence current demand. This subsection maps several leading models into the demand system (1), progressing from static demand to myopic, intertemporal portfolio choice, and models with deviations from full-information rational expectations (FIRE).

Static demand. The simplest case sets $S = 0$, so that no future expectations enter demand:

$$q_t = -A_0\, p_t + \xi_t. \tag{6}$$

The matrix A_0 contains the own- and cross-price elasticities, and there are no intertemporal links. This specification encompasses static demand systems used in the DSAP literature, including Koijen and Yogo (2019), Koijen et al. (2024), and Haddad et al. (2025a).[8]

Mean–variance (myopic) demand. Myopic mean–variance demand depends only on one-period-ahead conditional moments:

$$q_t = -A_0\, p_t + A_1\, \mu^p_{t,t+1} + B_1\, \mu^d_{t,t+1} + \xi_t, \tag{7}$$

where we hold conditional second moments constant for simplicity. Consider an investor with constant absolute risk aversion $\gamma > 0$ and perceived dollar return covariance matrix $\Sigma \succ 0$. Demand in levels is

$$Q_t = \frac{1}{\gamma}\, \Sigma^{-1} \big(\mathbb{E}_t[P_{t+1} + \mathcal{D}_{t+1}] - R^f\, P_t \big).$$

[8] Relative to Koijen and Yogo (2019), equation (6) permits a more flexible pattern of substitution. Appendix N provides the mapping.

Log-linearizing around $(\bar{Q},\bar{P},\bar{\mathcal{D}})$ delivers (7) with

$$A_0 = \mathrm{Diag}(\bar{Q})^{-1}\frac{R^f}{\gamma}\Sigma^{-1}\mathrm{Diag}(\bar{P}), \qquad A_1 = \beta A_0, \qquad B_1 = \mathrm{Diag}(\bar{Q})^{-1}\frac{1}{\gamma}\Sigma^{-1}\mathrm{Diag}(\bar{\mathcal{D}}), \tag{8}$$

where $\beta = 1/R^f$. The restriction $A_1 = \beta A_0$ arises because prices enter the portfolio problem through next-period expected returns.

Adjustment costs. Adjustment costs generate dependence on the entire expected path even when the investor's underlying objective is myopic. In Gârleanu and Pedersen (2013), the investor's position follows

$$Q_t - Q_{t-1} = \lambda_{tc}\big(\mathrm{Aim}_t - Q_{t-1}\big),$$

where $\lambda_{tc} \in (0,1)$ governs the speed of adjustment toward the aim portfolio. The aim portfolio is a geometrically weighted average of expected future myopic demands:

$$\mathrm{Aim}_t \propto \sum_{s=0}^{\infty} \psi_{tc}^{s}\,\Sigma^{-1}\,\mathbb{E}_t\big[P_{t+s+1} + \mathcal{D}_{t+s+1} - R^f P_{t+s}\big],$$

where $\psi_{tc} \in (0,1)$ denotes the decay rate. Log-linearization therefore generates geometrically declining sequence-space coefficients:

$$A_s \propto \psi_{tc}^{s-1}\,A_1, \qquad B_s \propto \psi_{tc}^{s-1}\,B_1, \qquad s = 1,2,\ldots,$$

where A_1 and B_1 are the myopic coefficients and ξ_t captures predetermined holdings. Adjustment costs thus generate nonzero coefficients at arbitrarily long horizons with a particular geometric restriction.

Intertemporal hedging demand. Merton (1973) shows that a long-horizon investor combines a myopic portfolio with a hedging portfolio that insures against adverse changes in future investment opportunities. Hedging demand is conventionally written as a function of the state variables that drive those opportunities. We now show that this state-space rule admits a sequence-space representation.

Consider one risky asset with constant return variance σ_r^2 and expected excess returns affine in an L-dimensional state vector X_t:

$$\mathbb{E}_t[r^e_{t+1}] = \mu_r + \psi_r^\top X_t, \qquad X_{t+1} = \Phi X_t + \nu_{t+1}.$$

The log-linear portfolio rule in Campbell and Viceira (2002) takes the form

$$\alpha_t = \underbrace{\frac{\mathbb{E}_t[r^e_{t+1}] + \sigma_r^2/2}{\gamma\sigma_r^2}}_{\text{myopic component}} + \underbrace{a_{\alpha,0} + a_{\alpha,1}^\top X_t}_{\text{hedging component}}. \tag{9}$$

The state dynamics imply

$$\mathbb{E}_t[r^e_{t+s}] - \mu_r = \psi_r^\top \Phi^{s-1} X_t, \qquad s \geq 1.$$

If expected returns over horizons $1, \ldots, S$ span the choice-relevant state component $a_{\alpha,1}^\top X_t$, there exist coefficients $\{\delta_s\}_{s=1}^S$ such that

$$a_{\alpha,1}^\top X_t = \sum_{s=1}^{S} \delta_s \big(\mathbb{E}_t[r^e_{t+s}] - \mu_r\big). \tag{10}$$

Substituting (10) into (9) expresses both the myopic and hedging components in terms of the

expected-return path. More generally, the joint paths of first and second moments entering (1) may span the relevant hedging component even when expected returns alone do not.

To isolate the intertemporal hedging channel, suppose that current prices affect quantities only through expected returns.[9] The resulting linearized demand takes the form

$$q_t = \sum_{s=1}^{S} D_s \mathbb{E}_t[r_{t+s}] + \xi_t.$$

Intertemporal hedging therefore naturally generates sensitivity to expected returns at longer horizons.

Behavioral and informational frictions. Models with behavioral beliefs can be accommodated when subjective forecasts admit a stable mapping into rational conditional forecasts (Gabaix, 2019, 2023; Bordalo et al., 2018). Systematic belief distortions then reshape the demand kernel, while residual belief shocks enter ξ_t. Appendix N.3 discusses in detail how to map behavioral and informational frictions into our dynamic demand system.

2.4 The econometric challenge

Our goal is to recover the dynamic demand kernel $\mathcal{K}$ from observed data. We assume that the econometrician observes a vector $\iota_t \in \mathbb{R}^N$ of instruments that shift outside demand, with one instrument for each asset. These instruments are orthogonal to the latent demand process ξ_t:

Assumption 1 (Latent demand exogeneity)**.** *The latent demand process satisfies*

$$\mathbb{E}[\xi_t \iota_{t-h}^\top] = \mathbf{0}_{N\times N} \qquad \textit{for all } h \geq 0.$$

[9] Appendix N allows for an additional direct price effect arising from the mapping of portfolio shares into quantities and from wealth.

Assumption 1 requires the instruments to be orthogonal not only to current latent demand, as in the static setting, but also to latent demand at all subsequent dates. This one-sided sequential exclusion restriction is necessary for dynamic recovery, but is weaker than the strict lead-lag exogeneity condition imposed on external proxies in local-projection instrumental-variables (LP-IV) estimation (Stock and Watson, 2018).

Recovering the dynamic demand kernel raises three identification challenges.

Challenge 1: instruments move multiple demand arguments. The structural demand coefficients answer ceteris paribus questions. For instance, the i-th diagonal element of A_0 gives the response of demand for asset i to its price, holding all other arguments of demand fixed. Even when $\iota_{i,t}$ satisfies the exogeneity condition, it need not generate this ceteris paribus variation. In the cross-sectional dimension, an outside-demand shock to one asset generally changes other asset prices when cross-price elasticities are nonzero (Fuchs et al., 2023). In the time-series dimension, a persistent outside-demand shock changes expected future prices and potentially other future arguments of demand (Binsbergen et al., 2025). Thus, an instrument that shifts outside demand generally moves several arguments of the demand system simultaneously.

Challenge 2: some demand arguments are latent. The conditional moments entering (1) are defined using investors' information and therefore need not be observed by the econometrician. Recovering the demand kernel must therefore proceed without specifying investors' state variables or directly observing the corresponding demand arguments.

Challenge 3: more coefficients, but no additional external instruments. The econometrician has access to N instruments that shift outside demand. This is the minimum number of linearly independent sources of variation needed to identify the unrestricted static demand system, which

contains N endogenous prices. The dynamic demand system introduces additional endogenous arguments and structural coefficients, but we do not assume access to additional external instruments. In particular, there is no separate external instrument for each expected future price, dividend, or risk. The usual simultaneous-equations intuition therefore suggests that identifying the richer dynamic system would require additional excluded instruments (He et al., 2025).

Section 3 studies the cross-price identification problem in a static setting. Section 4 then asks whether the same instruments can recover the dynamic demand kernel despite the additional latent arguments.

3 Recovering Structural Demand in Static Economies

In this section, we study recovery in a static economy with a rich pattern of cross-asset substitution. We show how the instruments introduced in Section 2.4 recover the contemporaneous structural demand matrix. To focus on the conditions for identification, we treat population projection coefficients as known and abstract from finite-sample issues.

3.1 The static economy

Equilibrium in the static economy. Consider the static special case of (1), in which A_s, B_s, and C_s are zero for all $s \geq 1$. Demand is

$$q_t = -A_0\, p_t + \xi_t. \tag{11}$$

Combining the demand system with the market-clearing condition (2), we can solve for the

equilibrium price:

$$p_t = A_0^{-1} \left[z_t + \xi_t \right].$$

This shows that equilibrium prices are correlated with the latent demand ξ_t. Hence, one needs an instrument to estimate the demand coefficients in Equation (11). Moreover, a shock that changes $z_{j,t}$ will generically change multiple prices. Individual instruments do not induce ceteris paribus variation in the price of a single asset.

The diagonal case. Consider the case where A_0 is a diagonal matrix, so cross-price effects are zero, but the diagonal elements are allowed to vary across assets. With heterogeneous own-price elasticities, one cannot recover the demand coefficients using a single cross-section. Time-series variation is necessary to identify the demand coefficients.[10]

Consider the following regression for asset j:

$$q_{j,t} = \beta_j p_{j,t} + u_{j,t}, \tag{12}$$

where we omit the constant term for simplicity.

Let $\iota_{j,t}$ be an instrument for asset j, that is, $\iota_{j,t}$ is orthogonal to the latent demand $\xi_{j,t}$ and it is correlated with outside demand $z_{j,t}$. Given the instrument, we can compute the first-stage and reduced-form responses:

$$p_{j,t} = \pi_{p,j} \iota_{j,t} + v^p_{j,t}, \qquad\qquad q_{j,t} = \pi_{q,j} \iota_{j,t} + v^q_{j,t}.$$

The coefficient $\pi_{p,j}$ corresponds to the *price multiplier* for asset j. It measures the impact of a

[10] Haddad et al. (2025a) make a similar point in the context of a static demand system with nonzero cross-elasticities.

unit change in the instrument on the price of asset j. The coefficient $\pi_{q,j}$ measures the response of residual supply to a unit change in the instrument.

Define the *instrumented elasticity* for asset j as

$$\epsilon_{j,0} \equiv -\frac{\pi_{q,j}}{\pi_{p,j}}. \tag{13}$$

This corresponds to minus the IV coefficient associated with regression (12). In the diagonal case, the price multiplier satisfies $\pi_{p,j} = -A_{0,jj}^{-1}\pi_{q,j}$, so $\epsilon_{j,0}$ recovers the own-price elasticity of demand. Provided $\pi_{q,j} \neq 0$, we can normalize the instrument so that $\pi_{q,j} = -1$. Under this normalization, the instrumented elasticity is simply the inverse of the price multiplier, $\epsilon_{j,0} = \pi_{p,j}^{-1}$. This shows how the price multiplier can be informative about the structural demand coefficients.

3.2 Static recovery

Cross-price multipliers. Consider next the case where the off-diagonal elements of A_0 are non-zero. In this case, $\iota_{j,t}$ is no longer a valid instrument for $p_{j,t}$ in regression (12). The residual $u_{j,t}$ absorbs the dependence of demand for asset j on the prices of other assets. Because the instrument induces variation in multiple prices at once, $u_{j,t}$ is correlated with $\iota_{j,t}$.

In this case, knowledge of the price multiplier $\pi_{p,j}$ is not enough to recover the demand coefficients. However, a generalization of the single-asset approach works. Consider the following first-stage and reduced-form multivariate regressions:

$$p_t = \Pi_p \iota_t + u_t^p, \qquad q_t = \Pi_q \iota_t + u_t^q.$$

We can interpret Π_p as a matrix of price multipliers. The diagonal elements capture the own-price multipliers, that is, they correspond to the equilibrium response of $p_{j,t}$ to a change in $\iota_{j,t}$. The

off-diagonal elements capture the *cross-price multipliers*, that is, the response of price $p_{j,t}$ to the instrument $\iota_{i,t}$, for $i \neq j$. The matrix Π_q characterizes the residual-supply responses.

The static recovery equation. Projecting both sides of (11) on ι_t and applying Assumption 1 gives the *static recovery equation*:

$$\Pi_q = -A_0 \Pi_p. \tag{14}$$

This is the central cross-equation restriction of the static economy: the same structural demand matrix must map the price response to the quantity response for every instrument.

Suppose the price responses satisfy the rank condition

$$\operatorname{rank}(\Pi_p) = N.$$

We can then define the *static instrumented elasticity* matrix as

$$\mathcal{E}_0 \equiv -\Pi_q \Pi_p^{-1}. \tag{15}$$

Equation (15) is the multi-asset analogue of equation (13), and it corresponds to a standard instrumental-variables estimator. Combining the recovery equation with the definition of instrumented elasticity, we obtain

$$\mathcal{E}_0 = A_0.$$

Thus, the instrumented elasticity equals the structural demand matrix in the static economy. Because A_0 is invertible, the rank condition also implies that Π_q is nonsingular. Since the scale and basis of the instruments are arbitrary, we can therefore rotate and rescale them so that $\Pi_q = -I_N$. Under

this normalization, the instrumented elasticity is simply the inverse of the price-multiplier matrix, $\mathcal{E}_0 = \Pi_p^{-1}$.

The standard conditions for identification. The combination of the market-clearing condition (2) and the demand system (11) corresponds to a standard linear simultaneous-equations model. Hence, identification requires the usual order and rank conditions to be satisfied (Koopmans, 1949). The order condition requires at least as many excluded sources of variation as endogenous prices. With N endogenous prices, we need at least N instruments. The rank condition requires those instruments to induce linearly independent price movements.

Each individual instrument may move every price. This corresponds to our first econometric challenge. However, what matters is that their responses collectively span the price space, so linear combinations of the observed equilibrium responses can reconstruct variation that changes one price while holding the others fixed.[11] This reconstructed ceteris paribus variation therefore identifies the columns of A_0.

In the static setting, identification combines a recovery equation, which summarizes the relevant cross-equation restrictions, with a rank condition ensuring that the instruments induce enough independent variation in prices. We show next that the dynamic identification problem has the same structure.

4 Recovering Structural Demand in Dynamic Economies

The static recovery result uses the equilibrium responses to the instruments to isolate A_0. When demand is forward-looking, the same instruments generally move current prices together with expected future prices, dividends, and risk. The impact instrumented elasticity therefore no longer

[11] See Fuchs et al. (2025) for a related discussion of the instruments required to identify a structural demand matrix.

isolates a structural demand coefficient. This section shows that the full paths generated by the same instruments can recover the dynamic demand kernel. Throughout, the relevant population projections are treated as known; their finite-sample estimation is taken up in Section 5.

4.1 The dynamic identification problem

We begin with a price-only benchmark in which A_s may be nonzero for $s \geq 1$, while $B_s = C_s = 0$. Appendix E extends the recovery result to the full demand system with nonzero dividend and risk coefficients.

The price process. Combining the price-only demand system with market clearing gives

$$A_0\, p_t - \sum_{s=1}^{S} A_s\, \mu^p_{t,t+s} = z_t + \xi_t. \tag{16}$$

Let $f_t \equiv z_t + \xi_t$ denote the forcing term. In the static case, $p_t = A_0^{-1} f_t$. In the dynamic case, current prices depend on the expected future path of f_t.

Define the conditional forward operator $F_t^s x_t \equiv \mathbb{E}_t[x_{t+s}]$. Then (16) can be written as

$$\mathcal{A}(F_t) p_t = f_t, \qquad \mathcal{A}(F_t) \equiv A_0 - \sum_{s=1}^{S} A_s F_t^s.$$

We assume that the demand system satisfies a standard invertibility condition.

Assumption 2 (Stable invertibility)**.** *The matrix polynomial* $\mathcal{A}(\lambda) = A_0 - \sum_{s=1}^{S} A_s \lambda^s$ *satisfies* $\det \mathcal{A}(\lambda) \neq 0$ *for every* $\lambda \in \mathbb{C}$ *such that* $|\lambda| \leq 1$.

Assumption 2 guarantees that the forward-looking demand operator has a unique stable one-sided inverse. We can use this inverse to solve for the equilibrium price process.

Proposition 1 (Equilibrium price process)**.** *Under Assumption 2, the inverse of $\mathcal{A}(F_t)$ admits the absolutely convergent expansion $\mathcal{A}(F_t)^{-1} = \sum_{\ell=0}^{\infty} \mathcal{A}_\ell F_t^\ell$, and the unique stable equilibrium price process is*

$$p_t = \sum_{\ell=0}^{\infty} \mathcal{A}_\ell \, \mathbb{E}_t[f_{t+\ell}],$$

where $\mathcal{A}_0 = A_0^{-1}$ and $\mathcal{A}_\ell = A_0^{-1} \sum_{j=1}^{\min(S,\ell)} A_j \mathcal{A}_{\ell-j}$ for $\ell \geq 1$.

The matrices $\mathcal{A}_\ell$ are equilibrium price kernels rather than primitive demand coefficients. They combine A_0 with the intertemporal coefficients A_s. Consequently, an impact price response generally reflects both structural demand and the expected path of the forcing process.

Intertemporal price multipliers. For the dynamic analysis, we residualize the instrument vector with respect to its joint history and retain the notation ι_t for the resulting innovation. This normalization does not change the contemporaneous instrument span used in Section 3. For any variable y_t, let Ψ_h^y denote its coefficient in the population projection on the instrument history. In particular, write

$$z_t = \sum_{h=0}^{\infty} \Psi_h^z \iota_{t-h} + \nu_{z,t}, \qquad p_t = \sum_{h=0}^{\infty} \Psi_h^p \iota_{t-h} + \nu_{p,t}, \tag{17}$$

where the residuals are orthogonal to the instrument history. The matrices Ψ_h^p correspond to *intertemporal price multipliers*, the dynamic counterparts of the price multiplier matrix in Section 3. Under Assumption 1, combining Proposition 1 with these projections gives

$$\Psi_h^p = \sum_{\ell=0}^{\infty} \mathcal{A}_\ell \Psi_{h+\ell}^z. \tag{18}$$

Thus, the intertemporal price multipliers combine the structural demand system with the path of outside demand induced by the instrument.

Projecting latent expectations. Let $\mathcal{G}_t$ denote the information generated by the instrument history and let $\mathcal{F}_t$ denote investors' information, with $\mathcal{G}_t \subseteq \mathcal{F}_t$. The econometrician need not observe $\mathbb{E}[p_{t+s} \mid \mathcal{F}_t]$. By iterated expectations,

$$\mathbb{E}[\mathbb{E}[p_{t+s} \mid \mathcal{F}_t] \mid \mathcal{G}_t] = \mathbb{E}[p_{t+s} \mid \mathcal{G}_t].$$

It follows that the coefficient from projecting investors' expectation of p_{t+s} on ι_{t-h} is Ψ^p_{h+s}. The same argument applies to expected dividends and any other conditional moment entering demand. This is the key projection result: latent expectations can be replaced by observable future response coefficients without specifying investors' state vector. It also does not require the history of observed economic variables to span investors' states or the complete vector of structural shocks.

Together, this projection property and the sequence-space representation overcome the second econometric challenge: investors' expectations need not be directly observed.

4.2 Recovery from intertemporal price multipliers

We now show how the intertemporal price multipliers recover the structural demand coefficients. The key is that the same coefficients must satisfy the equilibrium condition at every point along the instrument-induced response path.

The dynamic recovery equation. Projecting equilibrium condition (16) on each dated instrument and applying Assumption 1 gives a sequence of cross-equation restrictions in the spirit of Hansen and Sargent (1982).

Lemma 1 (Dynamic recovery equations)**.** *Suppose Assumption 1 holds. Then the structural demand*

coefficients $(A_0, \ldots, A_S)$ *and the price and outside-demand projection coefficients satisfy*

$$A_0 \Psi_h^p - \sum_{s=1}^{S} A_s \Psi_{h+s}^p = \Psi_h^z, \qquad h \geq 0. \tag{19}$$

Equation (19) is the dynamic analogue of the static recovery equation (14). Each dated value of the same primitive instrument contributes another equation involving the same structural coefficients. Because $q_t = -z_t$, the right-hand side is also minus the projection coefficient for elastic-investor holdings. Stacking these equations across horizons produces a system of restrictions whose identifying content depends on the variation in the intertemporal price multipliers.

Instrumented versus structural elasticities. The impact equation, corresponding to $h = 0$ in (19), is

$$A_0 \Psi_0^p - \sum_{s=1}^{S} A_s \Psi_s^p = \Psi_0^z.$$

Hence, the impact instrumented elasticity is

$$\mathcal{E}_0 = -\Pi_q \Pi_p^{-1} = \Psi_0^z (\Psi_0^p)^{-1} = A_0 - \sum_{s=1}^{S} A_s \Psi_s^p (\Psi_0^p)^{-1}.$$

Thus, the impact instrumented elasticity generally does not coincide with the contemporaneous demand coefficient A_0. Two important cases restore equality. First, demand may be static, so $A_s = \mathbf{0}_{N \times N}$ for all $s \geq 1$. Second, the instrument may induce a purely temporary price change, so $\Psi_s^p = \mathbf{0}_{N \times N}$ for all $s \geq 1$. When demand is dynamic and the instrument generates persistent price responses, even a full-rank impact response generally does not identify A_0. Recovering structural demand therefore requires using the remaining recovery equations rather than only the impact equation.

The finite MA case. To gain intuition, suppose the instrument affects outside demand at most $H \geq S$ periods ahead:

$$\Psi_h^z = \mathbf{0}_{N\times N} \quad \text{for all } h > H.$$

Equation (18) then implies $\Psi_h^p = \mathbf{0}_{N\times N}$ for all $h > H$. Evaluating the recovery equation (19) at $h = H$ gives the simplified condition:

$$A_0\Psi_H^p = \Psi_H^z.$$

Assuming that Ψ_H^p is full-rank, we can solve for A_0 as:

$$A_0 = \Psi_H^z(\Psi_H^p)^{-1}.$$

This is the static recovery equation evaluated at the end of the response path. The dated instrument ι_{t-H} moves p_t but no longer moves expected future prices. It therefore supplies the ceteris paribus variation needed to identify A_0.

Given A_0, we can recover A_1 by evaluating the recovery equation at $h = H - 1$:

$$A_0\Psi_{H-1}^p - A_1\Psi_H^p = \Psi_{H-1}^z$$

Solving for A_1, we get:

$$A_1 = \left[A_0\Psi_{H-1}^p - \Psi_{H-1}^z\right](\Psi_H^p)^{-1}.$$

Intuitively, $\iota_{t-(H-1)}$ moves both p_t and $\mathbb{E}_t[p_{t+1}]$. Together, ι_{t-H} and $\iota_{t-(H-1)}$ span the variation needed to separate A_0 from A_1.

The next proposition shows how to recover the entire demand system.

Proposition 2 (Finite MA recovery)**.** *Suppose the instrument affects outside demand at most $H \geq S$*

periods ahead and Ψ_H^p *is full-rank. Then, the structural demand coefficients* $(A_0, \ldots, A_S)$ *can be recovered from the instrument-induced response paths as follows:*

$$A_s = \left[A_0 \Psi^p_{H-s} - \sum_{j=1}^{s-1} A_j \Psi^p_{H+j-s} - \Psi^z_{H-s} \right] (\Psi^p_H)^{-1},$$

for $s = 1, \ldots, S$*, and* $A_0 = \Psi^z_H (\Psi^p_H)^{-1}$.

The finite-MA case is useful because it gives a constructive backward solution. The general identification result does not require an exact truncation.

The general price-only case. Stack the recovery equations for $h = 0, \ldots, H$ and define $\Psi^q_h \equiv -\Psi^z_h$.

Proposition 3 (The Recovery Theorem)**.** *Let the projections of prices and outside demand on the instrument history be given by* (17) *and suppose Assumption 1 holds. Stacking the recovery equations gives*

$$\overline{A}_{N \times N(S+1)} \overline{\Psi}^p_{N(S+1) \times N(H+1)} = \overline{\Psi}^q_{N \times N(H+1)},$$

where $\overline{A} = \left[-A_0, A_1, \cdots, A_S \right]$, $\overline{\Psi}^q = \left[\Psi^q_0, \Psi^q_1, \cdots, \Psi^q_H \right]$*, and*

$$\overline{\Psi}^p = \begin{bmatrix} \Psi^p_0 & \Psi^p_1 & \cdots & \Psi^p_H \\ \Psi^p_1 & \Psi^p_2 & \cdots & \Psi^p_{H+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi^p_S & \Psi^p_{S+1} & \cdots & \Psi^p_{S+H} \end{bmatrix}.$$

If the rank condition $\operatorname{rank}(\overline{\Psi}^p) = N(S+1)$ *is satisfied, then the stacked coefficient vector is given by:*

$$\overline{A} = \overline{\Psi}^q R_p,$$

where $R_p \equiv (\overline{\Psi}^p)^\top (\overline{\Psi}^p (\overline{\Psi}^p)^\top)^{-1}$.

The horizon count $H \geq S$ provides the necessary order condition. Identification additionally requires $\overline{\Psi}^p$ to contain $N(S+1)$ independent row directions, as formalized by the rank condition. Thus, dynamics generate additional identifying restrictions without requiring additional primitive external instruments.

How the Recovery Theorem addresses the three challenges. The Recovery Theorem shows how the three econometric challenges described in Section 2.4 can be overcome.

First, an instrument generally moves multiple arguments of demand simultaneously. In the price-only benchmark, it changes current prices together with expected future prices. By exploiting differences in these responses across assets and horizons, we can reconstruct the ceteris paribus variation needed to recover the structural coefficients, provided the rank condition holds.

Second, some arguments of demand may be latent. In particular, the econometrician need not observe investors' expectations or the state variables used to form them. The sequence-space representation and iterated expectations allow us to recover the relevant projections of these latent expectations from observable future response coefficients.

Third, dynamic demand contains more coefficients than static demand, but the econometrician does not receive additional primitive external instruments. The dynamic recovery equations provide additional cross-equation restrictions at different dates. For example, when $H = S = 1$, if $\mathbb{E}_t[p_{t+1}]$ were directly observed and the corresponding rank condition held, A_0 and A_1 could be recovered by a standard instrumental-variables estimator using $[\iota_t, \iota_{t-1}]$.[12] More generally, the propagation

[12] Because $\mathbb{E}_t[p_{t+1}]$ is typically unobserved, this standard instrumental-variables estimator is infeasible. The stacked recovery equations instead use iterated expectations to replace the relevant first-stage coefficients with observable future price responses.

of those innovations through equilibrium supplies the additional restrictions needed for dynamic recovery.

Persistence as a blessing, not a curse. A common criticism of static demand estimation is that the instruments typically induce persistent price responses. When demand is forward-looking, a static estimator then recovers a mixture of the contemporaneous and forward-looking demand coefficients. One proposed solution is to use instruments that induce purely transitory price variation (see, e.g., He et al., 2025). As shown above, the instrumented elasticity then coincides with the contemporaneous demand coefficient A_0. From this perspective, persistence appears to be the source of the problem.

However, counterfactuals generally require the entire dynamic demand kernel $\mathcal{K}$. If the instrument induces a purely transitory price response, so $\Psi_h^p = \mathbf{0}_{N\times N}$ for all $h \geq 1$, the rank condition cannot be satisfied when demand is dynamic ($S \geq 1$). In this case, we would not be able to recover the coefficients A_s for $s \geq 1$. In the absence of additional external instruments, persistence is therefore *necessary* for recovery of the dynamic demand system. In this sense, persistence is a blessing rather than a curse.

Persistence alone, however, does not guarantee that the rank condition is satisfied. Whether it does depends on whether the instrument-induced response path supplies enough independent variation across horizons. The next subsection characterizes this requirement in canonical time-series environments.

4.3 Persistence, dynamic rank, and identification

We now characterize the rank condition in more detail. To focus on the dynamic implications, we specialize to a single-asset economy, that is, $N = 1$. The Recovery Theorem implies that the $S + 1$ coefficients $(A_0, \ldots, A_S)$ are recoverable if and only if $\overline{\Psi}^p$ has rank $S + 1$.

The Hankel rank condition. To connect this finite-horizon condition to familiar time-series processes, define the *Hankel matrix* (see, e.g., Tsay, 2014) associated with a moving-average representation $\Psi = \{\Psi_h\}_{h\geq 0}$:

$$\mathcal{H}(\Psi) \equiv \begin{bmatrix} \Psi_0 & \Psi_1 & \Psi_2 & \cdots \\ \Psi_1 & \Psi_2 & \Psi_3 & \cdots \\ \Psi_2 & \Psi_3 & \Psi_4 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

The Hankel rank is the minimum number of independent dynamic components needed to generate the response path (Hannan and Deistler, 2012). A rank-one response follows a single propagation pattern, as in the AR(1) case. Higher rank means that multiple propagation patterns combine to generate independent variation across horizons. The recovery matrix $\overline{\Psi}^p$ is the finite block containing the first $S + 1$ rows and first $H + 1$ columns of $\mathcal{H}(\Psi^p)$. The rank of $\mathcal{H}(\Psi^p)$ therefore bounds the rank available to the finite recovery matrix.

The rank of $\mathcal{H}(\Psi^p)$ is, however, an endogenous object. To obtain a condition stated in terms of primitives, we therefore map this rank condition to properties of the outside-demand response induced by the instrument. The next lemma shows that the equilibrium mapping from outside demand into prices preserves Hankel rank.

Lemma 2 (Hankel-rank invariance)**.** *Suppose Assumptions 1 and 2 hold and* $\mathcal{H}(\Psi^z)$ *has finite rank* r*. Then* $\mathcal{H}(\Psi^p)$ *has the same rank:*

$$\operatorname{rank}\mathcal{H}(\Psi^p) = \operatorname{rank}\mathcal{H}(\Psi^z) = r.$$

Proof. See Appendix D.1. □

Lemma 2 shows that the Hankel rank for the price response is the same as the Hankel rank for the outside-demand response. Hence, the coefficients of the structural demand system determine the shape and magnitude of the price response but not its Hankel rank.

An important implication of Lemma 2 is that the rank condition in Proposition 3 will only be satisfied if $r \geq S+1$. If the Hankel rank for the outside-demand response is less than $S+1$, there is no horizon H that can satisfy the rank condition. Hence, the rank condition requires the instrument to induce a sufficiently rich outside-demand response.

The rank condition for autoregressive processes. Next, we show how the Hankel rank r is determined by the process for the projection of outside demand on the history of instrument innovations, $\hat{z}_t = \sum_{h=0}^{\infty} \Psi_h^z \iota_{t-h}$. Suppose $\hat{z}_t$ follows an autoregressive process with K distinct roots. The outside-demand response can then be written as

$$\Psi_h^z = \sum_{k=1}^{K} c_k \rho_k^h, \tag{20}$$

where the roots ρ_k are distinct and the loadings c_k are nonzero.[13]

In this case, the Hankel rank r is the number of distinct roots K. The next result shows that the rank condition in Proposition 3 can be satisfied if and only if $K \geq S+1$.

[13] In this case, the equilibrium price response is also a mixture of exponentials: $\Psi_h^p = \sum_{k=1}^{K} \frac{c_k}{\mathcal{A}(\rho_k)} \rho_k^h$.

Corollary 1 (Rank condition for AR processes)**.** *Suppose the outside-demand response is given by* (20)*, with K distinct roots and nonzero loadings, the conditions of Lemma 2 hold, and* $H \geq K - 1$*. Then the recovery matrix has rank*

$$\operatorname{rank} \overline{\Psi}^{p} = \min\{K, S+1\}.$$

A distinct-root AR(p*) response can therefore identify the full scalar demand kernel if and only if* $p \geq S + 1$*.*

Proof. See Appendix D.2. □

Corollary 1 gives a time-series interpretation of the rank condition: each distinct autoregressive root contributes an independent pattern of variation across horizons.[14]

Why the AR(1) case fails. A perhaps surprising implication of Corollary 1 is that the rank condition fails under the commonly assumed AR(1) process. For an AR(1) response, $\Psi^{z}_{h} = c\rho^{h}$ and invertibility implies $\Psi^{p}_{h} = c\rho^{h}/\mathcal{A}(\rho)$. Consequently, the recovery matrix factors as

$$\overline{\Psi}^{p} = \frac{c}{\mathcal{A}(\rho)} \begin{bmatrix} 1 \\ \rho \\ \vdots \\ \rho^{S} \end{bmatrix} \begin{bmatrix} 1 & \rho & \cdots & \rho^{H} \end{bmatrix}.$$

The matrix $\overline{\Psi}^{p}$ therefore has rank one for every H. Observing additional horizons supplies more recovery equations, but they are all proportional. The recovery equations identify only the

[14] Repeated roots generate responses such as $h\rho^{h}$, while complex-conjugate roots generate damped oscillations. These components also contribute to Hankel rank, even though they are not counted by the number of distinct real roots.

combination

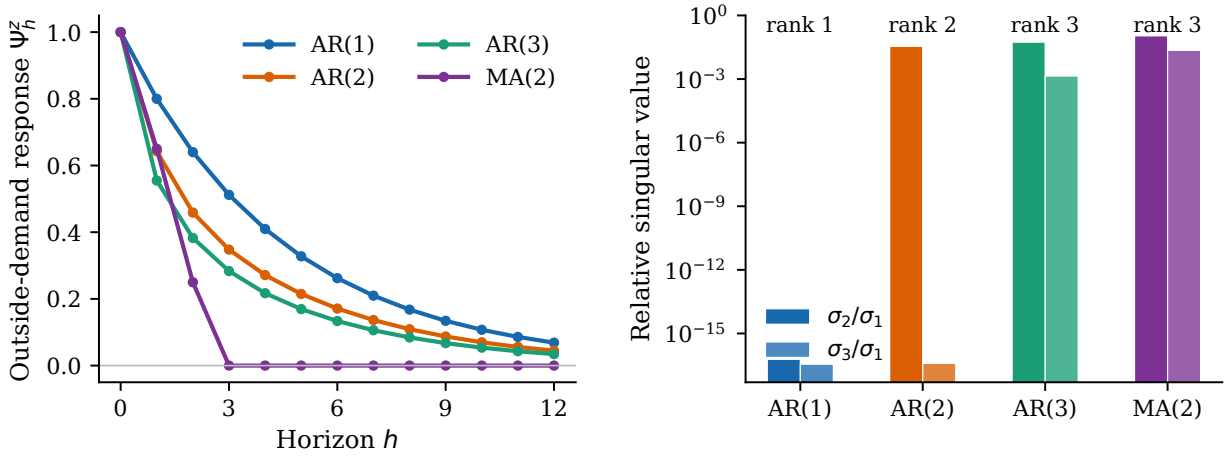
$$\mathcal{A}(\rho) = A_0 - \sum_{s=1}^{S} A_s \rho^s,$$

rather than the individual coefficients $(A_0, \ldots, A_S)$.

As illustrated by the AR(1) example in Section 2.2, investors' demand can then be written as a function of the current price alone, with coefficient $\mathcal{A}(\rho)$. A low value of $\mathcal{A}(\rho)$ can arise because demand is myopic and inelastic or because a more elastic contemporaneous response is offset by forward-looking demand. Because under AR(1) the instrument moves current and future prices in fixed proportions, these explanations are observationally equivalent. The AR(1) case therefore illustrates why persistence is necessary but not sufficient for identification.

Beyond AR(1). The rank failure in the AR(1) case is driven by its single propagation pattern. More flexible processes combine several propagation patterns and can therefore generate the independent variation needed for identification. For an autoregressive process with distinct roots, each additional root contributes another independent direction to the recovery matrix. Once the response contains at least $S+1$ such components, a sufficiently long recovery matrix has full row rank. More generally, the relevant count is the Hankel rank.

Figure 1 illustrates this progression for $S = 2$, so identification requires rank three. The AR(1) and AR(2) responses have ranks one and two and cannot identify all three structural coefficients. The AR(3) response reaches rank three and recovers the full demand kernel. The MA(2) response also has rank three, showing that the necessary variation can arise from a finite response distributed across several horizons. The right panel illustrates this rank condition using the singular values of the recovery matrix. Because matrix rank equals the number of nonzero singular values, the third singular values for the AR(3) and MA(2) responses are bounded away from zero, showing that both satisfy the rank condition.



(a) Response paths

(b) Relative singular values

Figure 1. Dynamic response paths and identification rank

Notes: The figure considers a scalar demand system with $S = 2$ and $(A_0,A_1,A_2) = (5,2.5,1)$. The left panel plots instrument-induced outside-demand responses for AR(1), AR(2), AR(3), and MA(2) specifications. The right panel reports the second and third singular values of the corresponding price recovery matrix, normalized by its largest singular value. The AR roots are (0.8), (0.8,0.35), and (0.8,0.45,0.1); the MA coefficients are (1,0.65,0.25).

These examples show that identification requires independent variation across horizons, which can arise from either multiple autoregressive propagation patterns or a finite response distributed across sufficiently many horizons. Appendix F provides an additional ARMA example illustrating the role of horizon selection.

Taken together, the results in this section show that persistence makes dynamic demand visible, while independent variation across the induced response path makes its coefficients separately recoverable.

5 Estimation and Empirical Implementation

Section 4 establishes recovery when the population response paths are known. We now replace those paths with estimates and adapt the recovery system to a large, unbalanced panel of stocks.

Estimation proceeds in two steps. We first estimate the response of each demand argument to the identified innovation at every required horizon. We then choose one structural demand kernel that jointly reconciles those estimated responses.

5.1 A two-step estimator

We consider again the general demand system (1). Combining the demand system with market clearing gives the condition:

$$z_t = A_0\, p_t - \sum_{s=1}^{S} A_s\, \mu^p_{t,t+s} - \sum_{s=1}^{S} B_s\, \mu^d_{t,t+s} + \sum_{s=1}^{S} C_s\, v^r_{t,t+s} - \xi_t,$$

Projecting both sides of the equation above onto the identified innovation at horizon h and using Assumption 1 gives the horizon-h recovery equation:

$$\Psi^z_h = A_0 \Psi^p_h - \sum_{s=1}^{S} A_s \Psi^p_{h+s} - \sum_{s=1}^{S} B_s \Psi^d_{h+s} + \sum_{s=1}^{S} C_s \Psi^v_{h+s}, \qquad h \geq 0, \tag{21}$$

Here, Ψ^d_h and Ψ^v_h denote the responses of realized dividends and realized second moments at horizon h, respectively. By iterated expectations, the responses of the corresponding date-t conditional arguments are Ψ^d_{h+s} and Ψ^v_{h+s}.

The left-hand side is the response of outside demand, while the right-hand side combines the responses of the price, dividend, and risk arguments of investor demand. Here, risk is a conditional second moment.[15] The structural coefficient matrices are the same at every response horizon h. This cross-equation restriction is the basis of estimation.

[15] Appendix E gives a more general formulation in which each horizon-specific risk object may have a separate response path.

A lower-dimensional parameterization. Let β collect the signed structural coefficients as they appear in the recovery equation: A_0 and $\{-A_s, -B_s, C_s\}_{s=1}^{S}$. The vector β will often be high-dimensional. However, motivated by economic models or considerations of estimation efficiency, we often consider a lower-dimensional parameterization of the structural coefficients. For instance, we show in Appendix N.1 that a linearized version of the demand system in Koijen and Yogo (2019) can be written as

$$q_{j,t} = \theta_1 p_{j,t} + \chi(\theta_1)^\top p_t + v_{j,t},$$

where $\chi(\theta_1)$ is a vector of coefficients. Hence, both the own- and cross-price coefficients, and therefore the entire $N \times N$ matrix A_0, are pinned down by the single parameter θ_1.

In our implementation, we consider an analogous dimensionality reduction. In particular, we allow the structural coefficients to be generated by a lower-dimensional parameter vector $\theta \in \Theta$,

$$\beta = g(\theta).$$

In the unrestricted case, $g(\theta) = \theta$. Economic models and factor structures can motivate restrictions on the admissible coefficient matrices that make θ lower-dimensional than β. These restrictions define the admissible demand systems and, by reducing the dimension of the parameter space, also mitigate overfitting and reduce the amount of independent response variation required for recovery. The next subsection describes the restrictions used in our stock-panel application.

First step: projection regressions. We perform the estimation in two steps. First, we estimate local projections (Jordà, 2005) of outside demand through horizon H and of the demand arguments through horizon $H + S$:

$$y_{t+h} = \alpha_h^y + \Psi_h^y \iota_t + u_{t+h}^y,$$

for $y \in \{z,p,d,v\}$, where α_h^y and Ψ_h^y are horizon-specific intercepts and response coefficients, respectively. All four projection regressions use realized outcomes rather than latent expectations. This produces estimated response coefficients $\widehat{\Psi}_h^y$. These regressions are the dynamic analogues of the reduced-form and first-stage regressions in an IV regression.

Second step: recovery across response horizons. In the second step, we recover the demand coefficients from the projection coefficients. We effectively treat the estimated projection coefficients as generated data and fit (21) across response horizons.

Intuitively, this step can be thought of as a regression in the generated data. The dependent variable is the estimated outside-demand response $\widehat{\Psi}_h^z$. The regressors are the current and shifted response paths of prices, dividends, and risk, $\widehat{\Psi}_h^p$ and $\{\widehat{\Psi}_{h+s}^p, \widehat{\Psi}_{h+s}^d, \widehat{\Psi}_{h+s}^v\}_{s=1}^S$. The coefficients are the structural demand matrices generated by $\beta = g(\theta)$.

The scalar price-only case makes the regression analogy transparent. With structural support S, the estimated recovery system is

$$\underbrace{\begin{bmatrix} \widehat{\Psi}_0^z \\ \widehat{\Psi}_1^z \\ \vdots \\ \widehat{\Psi}_H^z \end{bmatrix}}_{\widehat{Y}_H} \approx \underbrace{\begin{bmatrix} \widehat{\Psi}_0^p & \widehat{\Psi}_1^p & \cdots & \widehat{\Psi}_S^p \\ \widehat{\Psi}_1^p & \widehat{\Psi}_2^p & \cdots & \widehat{\Psi}_{S+1}^p \\ \vdots & \vdots & & \vdots \\ \widehat{\Psi}_H^p & \widehat{\Psi}_{H+1}^p & \cdots & \widehat{\Psi}_{H+S}^p \end{bmatrix}}_{\widehat{X}_H} \underbrace{\begin{bmatrix} A_0 \\ -A_1 \\ \vdots \\ -A_S \end{bmatrix}}_{\beta}.$$

The observations in this regression are response horizons rather than calendar dates. The equation at horizon h uses the outside-demand response at h and price responses through $h + S$ because current demand depends on expected future prices.

Notice that the matrix of regressors $\widehat{X}_H$ is the sample analogue of the transpose of $\overline{\Psi}^p$, the

key matrix in the Recovery Theorem. The corresponding population rank condition requires these columns to be linearly independent. The AR(1) failure in Section 4 is therefore the familiar problem of perfect multicollinearity among the generated regressors.

A similar logic applies to the general system. Vectorizing the matrix equations and stacking them through horizon H gives

$$Y_H(\mathbf{\Psi}) = X_H(\mathbf{\Psi})g(\theta), \qquad \varphi_H(\theta; \mathbf{\Psi}) \equiv Y_H(\mathbf{\Psi}) - X_H(\mathbf{\Psi})g(\theta), \tag{22}$$

where $\mathbf{\Psi}$ collects the population response paths. The vector $Y_H(\mathbf{\Psi})$ stacks the outside-demand responses, and the generated design matrix $X_H(\mathbf{\Psi})$ is constructed from the price, dividend, and risk responses. Appendix L gives the complete Kronecker-product construction.

Let $\widehat{Y}_H \equiv Y_H(\widehat{\mathbf{\Psi}})$ and $\widehat{X}_H \equiv X_H(\widehat{\mathbf{\Psi}})$ denote the generated outcome and design matrix constructed from the first-step estimates. We recover the structural parameters by minimum distance:

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \widehat{\varphi}_H(\theta)' W_H \widehat{\varphi}_H(\theta), \qquad \widehat{\varphi}_H(\theta) \equiv \widehat{Y}_H - \widehat{X}_H g(\theta). \tag{23}$$

If $g(\theta) = \theta$, the second step is (weighted) linear least squares. A lower-dimensional linear mapping gives restricted least squares, while a nonlinear mapping $g(\theta)$ corresponds to nonlinear least squares. The set Θ can additionally impose inequality restrictions on the structural coefficients.

The two-step estimator can also be represented as a stacked GMM system containing both the projection moments and the recovery restrictions. The two-step formulation is more transparent for exposition because it separates estimation of the response paths from structural recovery.

2SLS interpretation. The two-step estimator is closely related to two-stage least squares. If investors' expectations were directly observed and $g(\theta) = \theta$, one could estimate the structural

demand equation by two-stage least squares using $[\iota_t, \ldots, \iota_{t-H}]$ as instruments. The projection of outside demand on these instruments would be the reduced form, while the projections of the demand arguments would be the first stages. This estimator is infeasible because the expected demand arguments are latent. The sequence-space representation provides a feasible counterpart: by our projection result, the first-stage coefficients for the latent expectations can be recovered from projections of realized future outcomes on the same instruments. In a state-space representation, the analogous step would generally require the econometrician to observe or recover the relevant state variables. The second step relates these feasible first-stage coefficients to the reduced-form coefficients through the structural demand parameters. In the unrestricted linear case and with the corresponding weighting matrix, this procedure recovers the same coefficients as the infeasible two-stage least-squares benchmark. More generally, $g(\theta)$ allows us to impose economically motivated restrictions on the structural coefficients. Appendix J states the required relevance, exclusion, and lead–lag orthogonality conditions.

Inference with estimated projection coefficients. For point estimation, the second step treats $\widehat{Y}_H$ and $\widehat{X}_H$ as data. For inference, however, these objects cannot be treated as known. They are estimated from the same underlying observations, so their estimation errors are correlated across outcomes, response horizons, and the two sides of the recovery regression. Standard errors computed from the second-step regression alone would omit this first-step uncertainty.

We propagate this uncertainty with a moving-block bootstrap. In every draw, we resample the underlying observations, re-estimate all projection regressions, reconstruct $\widehat{Y}_H$ and $\widehat{X}_H$, and solve the structural recovery problem again. The resulting distribution of $\widehat{\theta}$ therefore incorporates uncertainty from both steps and the dependence induced by their common sample.

5.2 A lower-dimensional parameterization for stock panels

Our empirical application uses a large, unbalanced panel of stocks. Let N_t denote the number of stocks available at date t. Implementing the estimator requires two sets of high-dimensional objects. In the first step, each stock-specific innovation can affect every stock outcome, producing an $N_t \times N_t$ response matrix at each response horizon. In the second step, the structural demand kernel contains an $N_t \times N_t$ sensitivity matrix at each horizon. Estimating these matrices without restrictions is infeasible when N_t is large and the composition of the cross-section changes over time.

For the empirical implementation, we use the return-based representation in (3), which combines the price and dividend paths into expected returns. This representation is convenient because stock returns are typically modeled as stationary, whereas price levels are typically not. For the stocks available at date t, the demand system becomes

$$q_t = \sum_{s=1}^{S} D_{t,s}\mathbb{E}_t[r_{t+s}] - \sum_{s=1}^{S} C_{t,s}v^r_{t,t+s} + \xi_t,$$

where the date subscripts on $D_{t,s}$ and $C_{t,s}$ reflect changes in the composition of the stock cross-section.

Stock and portfolio representations. At date t, let $L_t \in \mathbb{R}^{N_t \times M}$ be an orthonormal matrix whose columns represent M portfolios, so $L_t'L_t = I_M$. Following the factor-plus-own decomposition of Haddad et al. (2025a), we parameterize the expected-return sensitivity at structural horizon s as

$$D_{t,s} = a_{D,s}I_{N_t} + L_t\Gamma^D_s L_t', \qquad D^F_s \equiv L_t'D_{t,s}L_t = a_{D,s}I_M + \Gamma^D_s.$$

The coefficients $a_{D,s}$ and Γ_s^D are common across dates, while L_t maps these coefficients into the stocks available at date t.

To see the mapping between stock and portfolio space, let $\mu_{t,s}^r \equiv \mathbb{E}_t[r_{t+s}]$ and decompose it as

$$\mu_{t,s}^r = L_t f_{t,s} + \mu_{t,s}^{r,\perp}, \qquad f_{t,s} = L_t' \mu_{t,s}^r, \qquad L_t' \mu_{t,s}^{r,\perp} = 0.$$

It follows that

$$D_{t,s}\mu_{t,s}^r = a_{D,s}\mu_{t,s}^{r,\perp} + L_t D_s^F f_{t,s}.$$

Thus, $a_{D,s}$ governs demand in stock-specific directions outside the portfolio span, while D_s^F governs demand and substitution among the M portfolios.

From one portfolio to five. The one-portfolio case makes the restriction transparent. For the equal-weight portfolio, $L_t = N_t^{-1/2}\mathbf{1}_{N_t}$, and

$$D_{t,s} = a_{D,s} I_{N_t} + \frac{\Gamma_s^D}{N_t}\mathbf{1}_{N_t}\mathbf{1}_{N_t}'.$$

Every cross-stock coefficient is Γ_s^D/N_t, while the own-stock coefficient is $a_{D,s} + \Gamma_s^D/N_t$. Our preferred specification generalizes this structure to $M = 5$ portfolios and sets structural support to $S = 4$. Estimating only in portfolio space would identify $D_s^F = a_{D,s} I_M + \Gamma_s^D$, but would not separate $a_{D,s}$ from the diagonal of Γ_s^D. Stock-level responses provide the variation outside the portfolio span that identifies $a_{D,s}$.

We impose the analogous structure on the first-step response matrices:

$$\Psi_{t,h}^z = a_z^{(h)} I_{N_t} + L_t G_z^{(h)} L_t', \qquad \Psi_{t,h}^r = a_r^{(h)} I_{N_t} + L_t G_r^{(h)} L_t'.$$

The first term captures the direct response of a stock to its own innovation, while the second captures cross-stock responses operating through the five-portfolio space. The response coefficients are common across dates, while L_t adapts the representation to the available cross-section.

Economic restrictions. At each structural horizon, we impose

$$a_{D,s} \geq 0, \qquad D_s^F = D_s^{F\prime} \succeq 0, \qquad s = 1, \ldots, 4. \tag{24}$$

Symmetry requires cross-factor responses to be reciprocal, while positive semidefiniteness ensures that an increase in expected returns along any factor-portfolio direction weakly raises demand in that direction. Both restrictions are invariant to orthogonal rotations of the factor basis. A general risk matrix would map all $M(M+1)/2$ factor second moments into all M factor-demand coordinates at every structural horizon. Let $\Sigma_{t,t+s}^F$ denote the conditional second-moment matrix of the portfolio returns and define $v_{t,t+s}^F \equiv \text{vech}(\Sigma_{t,t+s}^F)$. We restrict the stock-level response to risk according to

$$C_{t,s} v_{t,t+s}^r = L_t C_s^F v_{t,t+s}^F.$$

Thus, C_s^F maps portfolio-level risk into portfolio demand, and L_t maps that demand back into stock space. To limit the dimension of C_s^F, we retain only the five own-factor variance loadings:

$$C_s^F \,\text{vech}(\Sigma^F) = \begin{bmatrix} c_{1,s}\Sigma_{11}^F & \cdots & c_{5,s}\Sigma_{55}^F \end{bmatrix}', \qquad c_{i,s} \geq 0. \tag{25}$$

Risk enters demand with a minus sign, so $c_{i,s} \geq 0$ means that higher variance lowers demand for the same factor portfolio. The factor-plus-own form is a dimensionality restriction, whereas symmetry, positive semidefiniteness, and the risk signs are economic restrictions on the demand

kernel. Together, these restrictions define the lower-dimensional mapping $g(\theta)$ used in the second step.

5.3 Recovery and empirical implementation

We now specialize the two-step estimator in Section 5.1 to the stock-panel parameterization above. Let $\iota_{j,t}$ denote the identified innovation for stock j, let ι_t stack these innovations across stocks, and define the factor innovation $s_t = L_t'\iota_t$. For each horizon h and outcome $y \in \{z,r\}$, we estimate the pooled stock-week local projection

$$y_{j,t+h} = \alpha^y_{h,t} + \phi^{y\prime}_h \ell_{j,t} + a^{(h)}_y \iota_{j,t} + \ell'_{j,t} G^{(h)}_y s_t + u^y_{j,t+h}. \tag{26}$$

Date fixed effects absorb common weekly shocks, while predetermined loadings enter as controls. The coefficients $a^{(h)}_y$ and $G^{(h)}_y$ produce the fitted response matrix

$$\widehat{\Psi}^y_{t,h} = \widehat{a}^{(h)}_y I_{N_t} + L_t \widehat{G}^{(h)}_y L_t', \qquad y \in \{z,r\}.$$

We estimate the risk response directly in factor space. Let $f_d = L_t' r_d$ denote daily factor returns and let $\widetilde{\sigma}^F_{t,t+h}$ be the vectorized realized second moment in week $t+h$. For each horizon, we estimate

$$\widetilde{\sigma}^F_{t,t+h} = \alpha^\Omega_h + \Omega_h s_t + u^\Omega_{t+h}. \tag{27}$$

Estimating these projections separately at each horizon imposes neither a finite-order law of motion nor proportional response paths.

These estimated response paths are the generated data used in the second step described in Section 5.1. To accommodate the changing stock cross-section, at each sample week we construct

the stock-space recovery residual

$$\Delta_{h,t}(\theta) = \widehat{\Psi}^{z}_{t,h} + \sum_{s=1}^{S} D_{t,s}(\theta)\widehat{\Psi}^{r}_{t,h+s} - \sum_{s=1}^{S} L_t C^{F}_{s}(\theta)\widehat{\Omega}_{h+s}L_t'. \tag{28}$$

We retain its average diagonal and every entry of its factor compression:

$$e_{h,t}(\theta) = \begin{bmatrix} N_t^{-1}\operatorname{tr}(\Delta_{h,t}(\theta)) \\ \operatorname{vec}\left(\dfrac{M}{N_t}L_t'\Delta_{h,t}(\theta)L_t\right) \end{bmatrix} \in \mathbb{R}^{1+M^2}. \tag{29}$$

The average diagonal identifies the stock-specific component, while the factor compression requires the kernel to fit the full systematic response, including off-diagonal substitution. The scaling M/N_t puts a representative factor entry on the same stock-level scale as the average diagonal. We apply the fixed-weight minimum-distance estimator in (23) to these moments, subject to the economic restrictions in Section 5.2.

Each structural-horizon kernel is estimated separately. With $M = 5$, $S = 4$, and $H = 16$, the preferred system has

$$K_\theta = 4\left[1 + \frac{5(5+1)}{2} + 5\right] = 84$$

parameters and

$$K_m = (16+1)(1+5^2) = 442$$

recovery moments. The no-risk counterpart has 64 parameters.

These counts separate the order condition from the rank and strength conditions. Having more moments than parameters does not establish identification: the population recovery Jacobian must also have full column rank. Even under full rank, small singular values of the covariance-scaled sample Jacobian reveal weakly separated structural directions. A statistic for one return response

Algorithm 1 Estimating the dynamic demand kernel

1: Construct the identified stock-level innovations $\iota_{j,t}$.
2: Estimate the horizon-specific outside-demand and return responses using (26).
3: Estimate the horizon-specific factor second-moment responses using (27).
4: Construct the stock-specific and factor recovery moments in (29).
5: Apply the minimum-distance estimator in (23) subject to the restrictions in Section 5.2.
6: Evaluate the rank and conditioning of the recovery system.
7: Re-estimate both steps in moving-block bootstrap samples to conduct inference.

can establish that the instrument moves prices, but not that the entire response path identifies every dated coefficient.

Algorithm 1 summarizes the complete empirical procedure. Appendix A provides the sample-selection rules, numerical implementation, and exact bootstrap design.

The population rank condition determines whether the kernel is identified from the population response paths. The singular values of the estimated recovery system describe the stability of the inversion, while the bootstrap measures sampling uncertainty from both estimation steps. Section 6 reports these diagnostics together with the structural estimates.

6 Dynamic Demand in U.S. Stocks

6.1 Data, OFI innovation, and factor portfolios

We use weekly U.S. stock data from February 5, 1993, through December 30, 2022. OFI is constructed from WRDS intraday indicators using the Lee and Ready (1991) trade classification, aggregated to the stock-week level, and divided by lagged shares outstanding. A unit of OFI is therefore one share-equivalent of net buying as a fraction of shares outstanding. Weekly excess log returns subtract the weekly risk-free return compounded from the Ken French daily series. The final intersection contains 7,402,271 stock-weeks, 17,779 stocks, and 1,561 weeks.

Variable	Mean	SD	P25	P50	P75
Weekly OFI	-0.053	0.696	-0.207	-0.019	0.148
OFI innovation	0.001	0.619	-0.167	0.012	0.196
Weekly excess return	0.189	8.666	-3.425	-0.100	3.227

Table 1. Summary statistics for the recovery sample
This table reports stock-week summary statistics in percent. OFI is signed order-flow imbalance scaled by lagged shares outstanding. The OFI innovation is the residual from the within-stock forecasting regression described in the text.

The demand innovation is the residual from a pooled within-stock forecast of current OFI using four weekly OFI lags and four weekly excess-return lags. Specifically, we remove the within-stock means from OFI and all eight predictors, estimate the common forecasting coefficients, and winsorize the residual at the 0.5th and 99.5th percentiles. This residual is $\iota_{j,t}$. The exclusion restriction is that, conditional on this predetermined history, the residual moves returns, flows, and risk through outside demand rather than through the representative investor's latent demand shifter. Week fixed effects and predetermined factor loadings enter every stock-level response LP.

The preferred systematic basis uses 120 monthly characteristics from the Jensen-Kelly-Pedersen global factor data (Jensen, Kelly, and Pedersen, 2023); Appendix C lists every variable. The set spans accounting, return-history, and trading measures. For the 14 source fields supplied as accounting quantities divided by market equity, we first recover the underlying accounting quantity; neither contemporaneous price nor market equity is itself used as a factor weight. We lag every signal by one month, rank each characteristic within month into [0,1], subtract 0.5, and assign missing values the centered cross-sectional median of zero. Let K_t contain these signals and let $\widetilde{K}_t = [\iota_t \; K_t]$. We right-whiten the characteristic portfolios,

$$U_t = \widetilde{K}_t(\widetilde{K}_t'\widetilde{K}_t)^{-1/2}, \qquad U_t'U_t = I, \tag{30}$$

form daily managed returns $m_d = U_t' r_d$, and retain the first five principal components of the pooled managed-return covariance matrix. If $\widehat{V}_5$ contains those eigenvectors, the stock-level basis is

$$L_t = U_t \widehat{V}_5, \qquad L_t' L_t = I_5. \tag{31}$$

This construction fixes five empirical portfolios before structural recovery. It is the orthonormal characteristic-managed version of IPCA (Kelly, Pruitt, and Su, 2019). Appendix A reports the transparent equal-weight own/other basis as a separate check.

6.2 One shock generates rich identifying paths

Figure 2 connects the observed responses to the recovered demand system. OFI rises sharply on impact and then dissipates. The same innovation raises the impact return, but lowers the next week's own-stock return: the price increase creates a predictable reversal. The subsequent OFI and return responses vary across horizons rather than following a single proportional decay. These are the cross-equation movements that distinguish immediate demand from continuation demand.

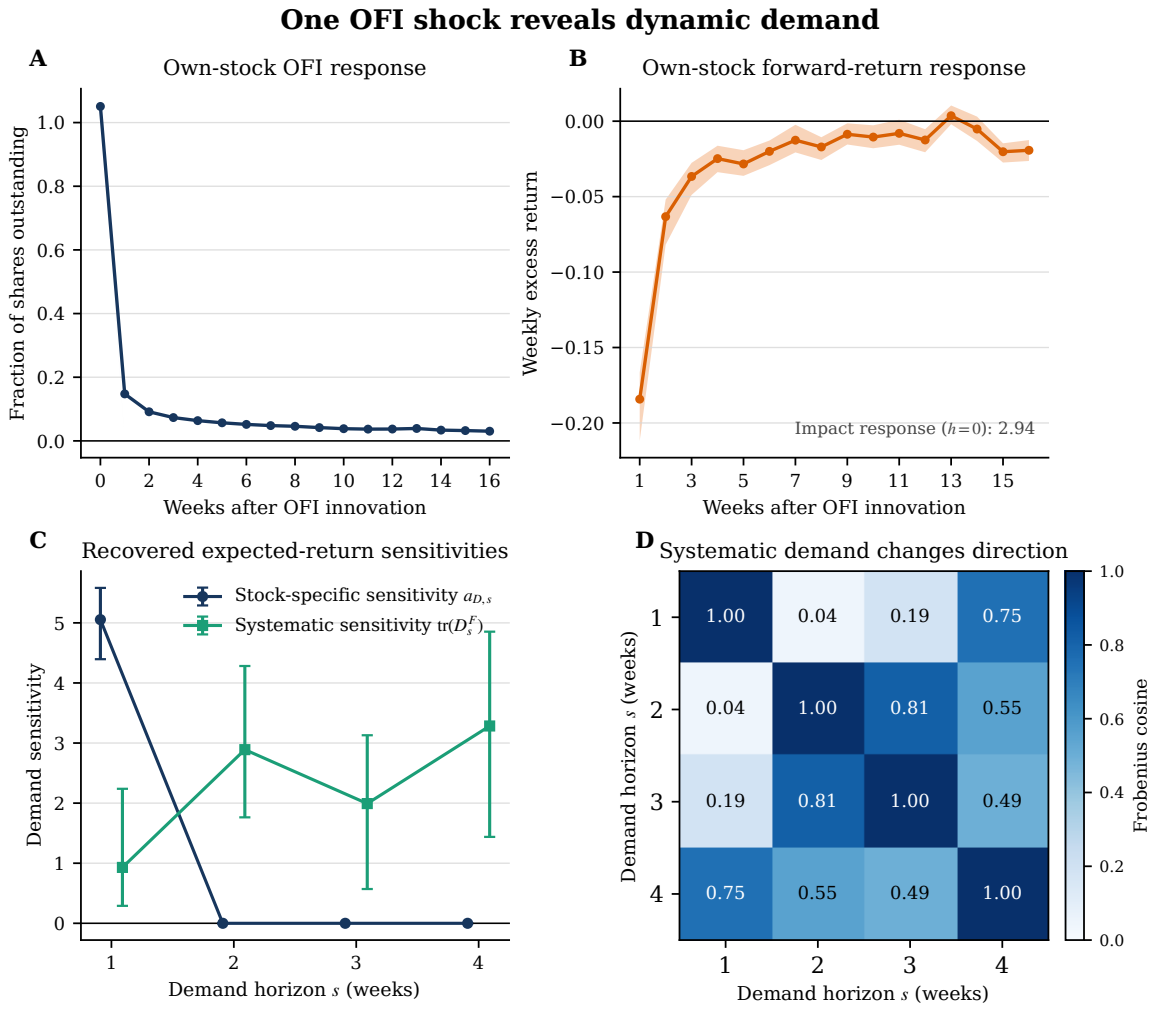


Figure 2. One OFI innovation reveals dynamic demand
Panels A and B plot the own-stock coefficients from horizon-specific OFI and forward one-week excess-return local projections. The shaded regions are 95% percentile intervals from 200 twelve-week block-bootstrap draws. The week-h return is the one-week return realized in that week, not a cumulative return. Panel C reports the stock-specific identity coefficient $a_{D,s}$ and the systematic trace $\mathrm{tr}(D_s^F)$ from the preferred four-week with-risk recovery, with 95% percentile intervals. Panel D reports pairwise Frobenius cosines of the four systematic expected-return matrices. The preferred recovery estimates 84 parameters from 442 moment conditions; all first-step response paths and structural parameters are re-estimated in each bootstrap draw, while the PCA basis is held fixed.

Instrument strength. The own-return response that anchors the roughly-five one-week elasticity is

$$\widehat{a}_r^{(1)} = -0.1842, \qquad \mathrm{SE}_{\mathrm{boot}} = 0.0113, \qquad F = t^2 = 267.3.$$

For a 5% test and a 10% modal-bias tolerance, Lewis and Mertens (2026) report critical values of 32.1 for one response and 64.1 for a four-response sequence. The OFI first stage exceeds both comfortably.

6.3 Structural elasticity

The economic scale is visible before solving the full system. A unit OFI innovation moves outside quantity by approximately one and lowers the following week's own return by about 0.18, so the demand sensitivity needed to absorb that quantity is of order $1/0.18$, or roughly five. The fixed-weight recovery then uses the complete own and factor response paths rather than this scalar approximation.

For comparison, we also report a one-week benchmark that relaxes the preferred cross-sectional restrictions in Section 5.2. This benchmark allows an unrestricted 5×5 expected-return block and, when risk is included, a full 5×15 covariance-risk operator. The preferred four-week specification instead imposes symmetry on the expected-return blocks and retains only the five own-factor variance loadings at each structural horizon.

Object	Estimate	Bootstrap SE	95% interval
Panel A. Structural own-price sensitivity			
Naive inverse multiplier	0.358	0.007	[0.344, 0.369]
One-week recovery, no risk	5.167	0.302	[4.600, 5.749]
One-week recovery, with covariance risk	5.188	0.302	[4.625, 5.771]
Four-week recovery, no risk	5.038	0.295	[4.385, 5.565]
Four-week recovery, with risk	5.049	0.296	[4.391, 5.576]
Panel B. Dynamic and risk effects			
Paired risk adjustment, one-week	0.021	0.005	[0.013, 0.030]
Paired risk adjustment, four-week	0.011	0.002	[0.006, 0.015]
Average own-variance loading, one-week	0.820	0.297	[0.332, 1.443]
Average own-variance loading, week 1	0.972	0.318	[0.102, 1.234]
Average own-variance loading, weeks 2–4	0.188	0.286	[0.000, 1.018]
Weeks 2–4 systematic expected-return sensitivity	8.167	1.330	[4.801, 10.010]

Table 2. Structural elasticity and dynamic demand
The naive inverse multiplier estimates one coefficient from one contemporaneous scalar equation. The one-week rows stack horizons $h = 0, \ldots, 4$, giving $5(1 + 5^2) = 130$ moments; the no-risk system estimates $1 + 5^2 = 26$ parameters and the covariance-risk system adds $5 \times [5(5 + 1)/2] = 75$ loadings, for 101 parameters. The four-week rows stack $h = 0, \ldots, 16$, giving $17(1 + 5^2) = 442$ moments; the no-risk system has $4(1 + 15) = 64$ parameters and the preferred with-risk system adds $4 \times 5 = 20$, for 84. Inequality restrictions do not reduce these counts. Panel B reports functions of the corresponding with-risk systems. Standard errors and percentile intervals use 200 twelve-week block-bootstrap draws that re-estimate every first-step LP and the constrained recovery. The combined weeks 2–4 risk loading is strictly positive in 97% of draws.

Table 2 shows a large and stable gap between the reduced-form multiplier and structural demand. The naive benchmark is 0.358, while the one-week and four-week recoveries all place the first-week stock-specific sensitivity near five. In the one-week covariance-risk system, controlling for risk

raises the average own sensitivity from 5.167 to 5.188; the paired increase is 0.021 with interval [0.013,0.030]. The average own-variance loading is positive, and the average stock-level cross sensitivity remains essentially zero at -0.00095. The main scale of the result therefore comes from the predictable return reversal, while the risk response moves the elasticity in the same economically required direction. Appendix I gives the scalar algebra and Appendix A reports every primitive one-week coefficient.

6.4 Dynamic demand

The bottom panels of Figure 2 show why the four-week estimator is not a static-demand exercise. The stock-specific identity component is concentrated in week one. Systematic demand, by contrast, remains active throughout the month: the traces of D_s^F are 0.930, 2.890, 1.993, and 3.284 in weeks one through four. Their combined weeks 2–4 sensitivity is 8.167, with a 95% interval of [4.801,10.010].

The systematic kernels also change direction across weeks. Their Frobenius cosines with the week-one matrix are 0.044, 0.186, and 0.754 at weeks two, three, and four. Continuation demand is therefore not a scalar decay of the first-week portfolio response. It reallocates demand across factor portfolios as the expected-return opportunity set evolves.

Risk has the expected sign in the preferred system. Controlling for own-factor variance raises the first-week own sensitivity from 5.038 to 5.049; the paired increase is 0.011 with interval [0.006,0.015]. The first-week average own-variance loading is 0.972, and the combined weeks 2–4 loading is 0.188 and positive in 97% of bootstrap draws. Thus the single OFI innovation identifies a strong immediate stock-specific response together with systematic expected-return and risk dynamics over the following month.

6.5 Focused robustness

Appendix A reports the one-factor own/other implementation, the complete five-factor one-week covariance-risk system, alternative risk restrictions, and the weekly primitive kernels. Across these designs, the recovered own sensitivity remains of order five and the average cross-stock sensitivity remains close to zero, consistent with the small cross-price effects documented by Chaudhry and Davis (2026).

Appendix K addresses direct dependence on past quantities, as emphasized by Van der Beck (2025). Adding a common lagged-quantity coefficient uses no additional instrument: the coefficient is recovered from the current and lagged quantity responses generated by the same OFI innovation. The estimate selects zero in every bootstrap draw, leaving the first-week and continuation-demand results unchanged. Adding four predetermined OFI lags and four predetermined return lags directly to every stock-level first-step projection likewise preserves a first-week own sensitivity near five and a precisely positive weeks 2–4 systematic trace. The dynamic result is therefore not generated by omitted pre-shock trading history.

7 Conclusion

This paper studies how to recover and use demand elasticities in dynamic asset markets. Because a price movement today must be accompanied by some path of future expected dividends, returns, or risk, there is no single context-free elasticity. The object recovered by an empirical design depends on the process generating the price movement. This is not a defect of demand-system asset pricing. It is the accounting structure of dynamic asset markets.

We formalize this point by distinguishing structural elasticities from instrumented elasticities. The structural elasticity is the demand response holding fixed the other arguments of the demand

system. The instrumented elasticity is the response induced by a particular shock, including the way that shock moves future expected returns and risk. Under linearized market clearing, the inverse of the instrumented elasticity is the price multiplier for the identifying shock. The multiplier is therefore a valid reduced-form object, but it is not generally a transportable structural parameter.

The recovery result shows how to move from the reduced-form object back to structural demand. The path of flow, return, and risk impulse responses imposes a system of restrictions on the structural coefficients. For stock data, we implement this idea with an orthonormal factor representation that accommodates unbalanced panels and keeps the recovery problem low-dimensional. The stock-level local projections identify the own identity component of demand, while the factor projection captures common substitution and risk channels.

In our data, the quantitative gap is large. The naive inverse-multiplier benchmark is about 0.35, while the recovered stock-specific own elasticity is about 5. The dynamic recovery reveals a distinction hidden by this scalar summary. The stock-specific component is concentrated in week one, but systematic expected-return sensitivity is positive and precisely estimated in every week through week four, with the systematic matrices rotating across factor directions. Controlling for own-factor variance raises the first-week elasticity by a precisely estimated amount, and the combined risk loading beyond week one is positive in 97% of bootstrap draws. Thus the same OFI shock identifies strong immediate own-price demand and economically important expected-return and risk dynamics without requiring additional instruments.

References

Albuquerque, Rui, Martin Eichenbaum, Victor Xi Luo, and Sergio Rebelo, 2016, Valuation risk and asset pricing, *The Journal of Finance* 71, 2861–2904.

Baba Yara, Fahiz, Brian H. Boyer, and Carter Davis, 2021, The limits of factor model spanning, SSRN working paper, Available at SSRN.

Bianchi, Francesco, Cosmin Ilut, and Hikaru Saijo, 2024, Diagnostic business cycles, *The Review of Economic Studies* 91, 129–162.

Binsbergen, Jules H van, Benjamin David, and Christian C Opp, 2025, How (not) to identify demand elasticities in dynamic asset markets, Working paper.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2018, Diagnostic expectations and credit cycles, *The Journal of Finance* 73, 199–227.

Campbell, John Y, and Luis M Viceira, 2002, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors* (Oxford University Press).

Chaudhry, Aditya, and Carter Davis, 2026, The origins of the factor zoo: Investors weakly substitute across stocks, Working paper.

Cox, John C, and Chi-fu Huang, 1989, Optimal consumption and portfolio policies when asset prices follow a diffusion process, *Journal of Economic Theory* 49, 33–83.

Davis, Carter, Mahyar Kargar, and Jiacui Li, 2025, Why do portfolio choice models predict inelastic demand?, *Journal of Financial Economics* 172, 104096.

Davis, Carter, Samuli Knüpfer, Jens Soerlie Kvaerner, Bahar Sen Dogan, and Petra Vokata, 2024, Do households matter for asset prices?, Working paper.

Fuchs, William, Satoshi Fukuda, and Daniel Neuhann, 2023, Demand-system asset pricing: Theoretical foundations, Working paper.

Fuchs, William, Satoshi Fukuda, and Daniel Neuhann, 2025, A trilemma for asset demand estimation, Working paper.

Gabaix, Xavier, 2019, Behavioral inattention, in *Handbook of Behavioral Economics: Applications and Foundations 2*, volume 2, 261–343 (Elsevier).

Gabaix, Xavier, 2023, Behavioral macroeconomics via sparse dynamic programming, *Journal of the European Economic Association* 21, 2327–2376.

Gârleanu, Nicolae, and Lasse Heje Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 68, 2309–2340.

Grossman, Sanford J, and Joseph E Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American Economic Review* 70, 393–408.

Haavelmo, Trygve, 1943, The statistical implications of a system of simultaneous equations, *Econometrica* 11, 1–12.

Haddad, Valentin, Zhiguo He, Paul Huebner, Péter Kondor, and Erik Loualiche, 2025a, Causal inference for asset pricing, Working paper.

Haddad, Valentin, Paul Huebner, and Erik Loualiche, 2025b, How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing, *American Economic Review* 115, 975–1018.

Hannan, Edward J., and Manfred Deistler, 2012, *The Statistical Theory of Linear Systems*, volume 70 of *Classics in Applied Mathematics* (Society for Industrial and Applied Mathematics).

Hansen, Lars Peter, and Thomas J Sargent, 1982, Instrumental variables procedures for estimating linear rational expectations models, *Journal of Monetary Economics* 9, 263–296.

He, Zhiguo, Péter Kondor, and Jessica S Li, 2025, Demand elasticity in dynamic asset pricing, Working paper.

Heckman, James J, and Rodrigo Pinto, 2024, Econometric causality: The central role of thought experiments, *Journal of Econometrics* 243, 105719.

Hurwicz, Leonid, 1962, On the structural form of interdependent systems, in Ernest Nagel, Patrick Suppes, and Alfred Tarski, eds., *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, 232–239 (Stanford University Press, Stanford).

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *Journal of Finance* 78, 2465–2518.

Jordà, Òscar, 2005, Estimation and inference of impulse responses by local projections, *American Economic Review* 95, 161–182.

Karatzas, Ioannis, John P Lehoczky, and Steven E Shreve, 1987, Optimal portfolio and consumption decisions for a "small investor" on a finite horizon, *SIAM Journal on Control and Optimization* 25, 1557–1586.

Kelly, Bryan T, Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

Koijen, Ralph SJ, Robert J Richmond, and Motohiro Yogo, 2024, Which investors matter for equity valuations and expected returns?, *Review of Economic Studies* 91, 2387–2424.

Koijen, Ralph SJ, and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Koopmans, Tjalling C, 1949, Identification problems in economic model construction, *Econometrica* 17, 125–144.

Lee, Charles MC, and Mark J Ready, 1991, Inferring trade direction from intraday data, *The Journal of Finance* 46, 733–746.

Lewis, Daniel J., and Karel Mertens, 2026, Weak instrument bias in impulse response estimators, Working Paper 2601, Federal Reserve Bank of Dallas.

Ljungqvist, Lars, and Thomas J Sargent, 2018, *Recursive Macroeconomic Theory*, fourth edition (MIT Press).

Mas-Colell, Andreu, Michael D Whinston, and Jerry R Green, 1995, *Microeconomic Theory* (Oxford University Press).

Merton, Robert C, 1969, Lifetime portfolio selection under uncertainty: The continuous-time case, *The Review of Economics and Statistics* 51, 247–257.

Merton, Robert C, 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Plagborg-Møller, Mikkel, and Christian K Wolf, 2021, Local projections and VARs estimate the same impulse responses, *Econometrica* 89, 955–980.

Stock, James H, and Mark W Watson, 2018, Identification and estimation of dynamic causal effects in macroeconomics using external instruments, *The Economic Journal* 128, 917–948.

Tamoni, Andrea, Stanislav Sokolinski, and Yizhang Li, 2024, Which investors drive anomaly returns and how?, Working paper.

Tsay, Ruey S., 2014, *Multivariate Time Series Analysis: With R and Financial Applications* (John Wiley & Sons).

Van der Beck, Philippe, 2021, Flow-driven ESG returns, Technical report, Swiss Finance Institute.

Van der Beck, Philippe, 2025, Short versus long-run demand elasticities in asset pricing, Harvard Business School Working Paper 26-070.

Vayanos, Dimitri, and Jean-Luc Vila, 2021, A preferred-habitat model of the term structure of interest rates, *Econometrica* 89, 77–112.

Appendix

A Empirical Construction and Supporting Estimates

This appendix supplies the exhibits and construction details behind the consolidated evidence in Section 6. All variants use the same stock-level OFI innovation and the same horizon-by-horizon local-projection design. They differ only in the factor basis, structural support, and economic restrictions stated below.

We keep weeks with at least 25 usable stocks in the pooled stock-level local projections. The return outcome at horizon h is the one-week excess return realized in week $t + h$, rather than the cumulative return through that horizon. Before structural recovery, we reshape each column of $\widehat{\Omega}_h$ into a symmetric factor second-moment response and project it onto the negative-semidefinite cone. We repeat this projection in every bootstrap draw. This first-step restriction says that the OFI innovation weakly lowers future factor risk in the estimated shock direction and is distinct from the sign restrictions on investors' demand response to risk. The structural minimum-distance problem is a convex quadratic program before the cone restrictions, which we enforce using supporting hyperplanes until every restriction is satisfied.

For exact bootstrap replication, blocks are sampled as circular twelve-week moving blocks. The preferred four-week estimates use 200 draws with seed 12345. The one-step covariance-risk results combine four 50-draw streams with seeds 86420, 97531, 108642, and 119753. The standalone 200-draw response bands in Figure 2 use seed 54321. Every draw re-estimates the relevant first-stage projections and structural recovery; the PCA basis is held fixed.

A.1 One-factor own/other implementation

The equal-weight basis is

$$L_t = \ell_t = \frac{1}{\sqrt{N_t}}\iota_t, \qquad \ell_t'\ell_t = 1.$$

The expected-return sensitivity becomes

$$D_{t,s} = a_{D,s} I_{N_t} + \frac{\Gamma_s^D}{N_t}\iota_t\iota_t'. \tag{A.1}$$

LP object	Intercept	Avg. diag. response	Avg. off-diag. response
Returns LP, $h = 0$	3.089 (0.063)	-0.001 (0.003)	>-0.0001 (<0.0001)
Returns LP, $h = 1$	-0.169 (0.010)	>-0.0001 (0.001)	>-0.0001 (<0.0001)
Flows LP, $h = 0$	1.053 (0.001)	-0.001 (0.001)	>-0.0001 (<0.0001)
Risk LP, $h = 1$	3.060 (0.182)	-26.161 (4.204)	

Table A.1. First-stage response estimates in the one-factor own/other basis
The table summarizes the stock-level OFI and forward-return local projections and the factor-level risk projection. "Intercept" is the own-stock coefficient $a^{(h)}$ for returns and flows and the baseline factor-variance intercept for risk. Average diagonal and off-diagonal entries summarize the factor interaction. Each horizon is estimated separately using the stock-week regression in (26).

Specification	Avg. diag. D_1	Avg. off-diag. D_1
Naive IV benchmark	0.341 (0.007)	
One-step, no risk	5.520 (0.397)	-0.00762 (0.00663)
One-step, with risk	6.741 (0.460)	1.15437 (0.39042)

Table A.2. Structural recovery in the one-factor own/other basis
The naive row uses one contemporaneous scalar equation for one coefficient. The one-step rows stack $h = 0, \ldots, 4$, giving $5(1 + 1^2) = 10$ moments. The no-risk system has $1 + 1^2 = 2$ parameters; adding the scalar variance loading gives three. Bootstrap standard errors re-estimate both stages in 200 twelve-week block draws.

Its diagonal is $a_{D,s} + \Gamma_s^D / N_t$ and every off-diagonal entry is Γ_s^D / N_t. Thus $a_{D,s}$ is the own-minus-other component and Γ_s^D / N_t is the common cross-stock component. The risk state is the weekly sum of squared daily equal-weight factor returns. The one-step specification imposes

$$D_s^F = C_s^F = 0 \qquad \text{for all } s \geq 2.$$

Parameter	No risk	With risk	Parameter	No risk	With risk
$a_{D,1}$	5.528 (0.399)	5.515 (0.398)	Γ_1^D	-39.731 (34.564)	6020.001 (2036.031)

Table A.3. Primitive D_1 parameters in the one-factor own/other basis
The table reports $a_{D,1}$ and Γ_1^D from (A.1). Bootstrap standard errors are in parentheses. These are the same two-parameter no-risk and three-parameter with-risk systems estimated from 10 moments in Table A.2.

Specification	Factor-variance C_1^F
One-step, with risk	30.218 (11.199)

Table A.4. Risk loading in the one-factor own/other basis
The table reports the scalar loading of factor demand on the equal-weight factor variance. It is estimated jointly with the two expected-return coefficients, for three parameters and 10 moments.

A.2 Five-factor one-step implementation

The five-factor basis follows (30)–(31). Daily factor returns are $f_d = L_t' r_d$. At horizon h, the risk outcome is

$$\widetilde{\Sigma}^F_{t,t+h} = \sum_{d \in \mathcal{W}_{t+h}} f_d f_d', \qquad \widetilde{\sigma}^F_{t,t+h} = \mathrm{vech}(\widetilde{\Sigma}^F_{t,t+h}).$$

It has 15 entries and is a weekly realized second moment rather than a demeaned sample covariance. The estimated 15×5 response $\widehat{\Omega}_h$ maps each factor OFI shock into all factor variance and covariance states.

The one-step covariance-risk benchmark allows a full 5×15 risk operator. For factor-demand row i, write its action on a covariance matrix as

$$d_i' \,\mathrm{vech}(\Sigma) = \mathrm{tr}(\mathcal{B}_i \Sigma), \qquad \mathcal{B}_i \succeq 0. \tag{A.2}$$

The five row-PSD restrictions mean that higher covariance risk in any factor direction weakly lowers demand. This property is invariant to sign changes and rotations of the factor portfolios, even though individual covariance-coordinate coefficients are not.

LP object	Intercept	Avg. diag. response	Avg. off-diag. response
Returns LP, $h = 0$	2.937	13.422	1.501
	(0.055)	(0.354)	(0.354)
Returns LP, $h = 1$	-0.184	0.677	0.421
	(0.011)	(0.270)	(0.251)
Flows LP, $h = 0$	1.051	0.116	-0.015
	(0.001)	(0.010)	(0.006)
Risk LP, $h = 1$	0.880	-7.449	-0.627
	(0.050)	(1.404)	(0.207)

Table A.5. First-stage response estimates in the five-factor basis
The table summarizes the stock-level return and OFI local projections and the factor-level covariance-risk projection. "Intercept" is the own-stock coefficient $a^{(h)}$ for returns and flows and the average diagonal baseline second moment for risk. Risk-response columns summarize the horizon-one 15×5 response matrix. Each horizon is estimated separately.

Specification	Avg. diag. D_1	Avg. off-diag. D_1
Naive IV benchmark	0.358	
	(0.007)	
One-step, no risk	5.167	-0.00096
	(0.302)	(0.00007)
One-step, with risk	5.188	-0.00095
	(0.302)	(0.00008)

Table A.6. Structural recovery in the five-factor basis
The naive row uses one contemporaneous scalar equation for one coefficient. The one-step rows stack $h = 0, \ldots, 4$, giving $5(1 + 5^2) = 130$ moments. The no-risk model estimates $1 + 5^2 = 26$ parameters. The row-PSD covariance-risk model adds $5 \times [5(5 + 1)/2] = 75$ loadings, for 101 parameters. Inequality restrictions do not reduce the count; the sample Jacobian has rank 101.

Parameter	No risk	With risk	Parameter	No risk	With risk
$a_{D,1}$	5.172 (0.302)	5.194 (0.302)	$(\Gamma_1^D)_{3,3}$	-4.241 (0.787)	-4.842 (0.810)
$(\Gamma_1^D)_{1,1}$	-5.379 (0.327)	-5.119 (0.373)	$(\Gamma_1^D)_{3,4}$	4.654 (1.425)	4.040 (1.355)
$(\Gamma_1^D)_{1,2}$	0.487 (0.317)	0.859 (0.429)	$(\Gamma_1^D)_{3,5}$	-1.050 (2.377)	-2.424 (2.204)
$(\Gamma_1^D)_{1,3}$	-0.153 (0.314)	-0.456 (0.394)	$(\Gamma_1^D)_{4,1}$	-0.358 (0.275)	-0.019 (0.317)
$(\Gamma_1^D)_{1,4}$	1.688 (0.556)	1.198 (0.800)	$(\Gamma_1^D)_{4,2}$	2.057 (0.972)	2.527 (0.871)
$(\Gamma_1^D)_{1,5}$	-0.923 (1.062)	-1.794 (1.477)	$(\Gamma_1^D)_{4,3}$	-0.656 (1.036)	-1.453 (0.853)
$(\Gamma_1^D)_{2,1}$	0.770 (0.226)	1.189 (0.281)	$(\Gamma_1^D)_{4,4}$	-5.351 (1.673)	-6.248 (1.456)
$(\Gamma_1^D)_{2,2}$	-6.736 (0.719)	-6.191 (0.823)	$(\Gamma_1^D)_{4,5}$	1.448 (2.481)	0.409 (2.065)
$(\Gamma_1^D)_{2,3}$	-1.927 (0.624)	-2.234 (0.800)	$(\Gamma_1^D)_{5,1}$	-0.641 (0.329)	-0.341 (0.355)
$(\Gamma_1^D)_{2,4}$	-0.638 (0.939)	-1.573 (1.001)	$(\Gamma_1^D)_{5,2}$	-0.019 (0.987)	0.467 (0.921)
$(\Gamma_1^D)_{2,5}$	2.816 (2.167)	0.897 (2.420)	$(\Gamma_1^D)_{5,3}$	-0.574 (1.016)	-0.306 (0.961)
$(\Gamma_1^D)_{3,1}$	-1.271 (0.287)	-0.824 (0.296)	$(\Gamma_1^D)_{5,4}$	4.349 (1.536)	2.907 (1.423)
$(\Gamma_1^D)_{3,2}$	0.837 (0.776)	1.466 (0.730)	$(\Gamma_1^D)_{5,5}$	-3.672 (3.377)	-6.209 (3.346)

Table A.7. Primitive D_1 parameters in the five-factor basis
The table reports the coefficients in $D_{t,1} = a_{D,1} I_{N_t} + L_t \Gamma_1^D L_t'$. Bootstrap standard errors are in parentheses. The no-risk columns contain 26 parameters and the covariance-risk columns are part of the 101-parameter system; both use 130 moments.

Specification	Avg. own-variance C_1	Avg. other-covariance C_1
One-step, with risk	0.820	0.262
	(0.297)	(0.114)

Table A.8. Risk loading in the five-factor basis
The own-variance column averages the loadings of factor-i demand on factor-i variance. The other-covariance column averages the remaining entries of the 5×15 operator. These 75 coefficients are jointly estimated with 26 expected-return coefficients using 130 moments.

C restriction	No-risk avg. diag. D_1	With-risk avg. diag. D_1
Own variance nonnegative	5.167	5.164
	(0.302)	(0.301)
Row-PSD risk kernel	5.167	5.188
	(0.302)	(0.302)

Table A.9. Five-factor recovery under alternative risk-loading restrictions
The table holds the first-stage paths fixed and compares restrictions on the one-step 5×15 risk operator. Each with-risk row jointly estimates 26 expected-return and 75 risk coefficients from 130 moments. The row-PSD model imposes (A.2); the own-variance comparison constrains only five named variance coordinates.

A.3 Weekly structural kernels

The preferred four-week estimator imposes (24)–(25) and otherwise estimates every weekly matrix separately. The following tables report the weekly primitives and the functionals consolidated in Table 2.

	Week 1	Week 2	Week 3	Week 4
Identity component $a_{D,s}$	5.054	0.000	0.000	0.000
	(0.296)	(0.000)	(0.000)	(0.000)
Avg. stock own sensitivity $\overline{D}_{s,ii}$	5.04917	0.00059	0.00041	0.00066
	(0.29556)	(0.00013)	(0.00013)	(0.00018)
Systematic sensitivity $\mathrm{tr}(D_s^F)$	0.930	2.890	1.993	3.284
	(0.533)	(0.655)	(0.642)	(0.877)
Avg. own-variance loading $\overline{c}_s$	0.972	0.074	0.113	0.000
	(0.318)	(0.099)	(0.286)	(0.035)

Table A.10. Four-week demand kernels
The identity component $a_{D,s}$ acts on stock directions outside the factor span. Average stock own sensitivity is the average diagonal of $D_{t,s}$, systematic sensitivity is $\mathrm{tr}(D_s^F)$, and the final row averages the five own-variance loadings. The estimator has $4(1 + 15 + 5) = 84$ parameters and uses $17(1 + 5^2) = 442$ moments through $h = 16$. Bootstrap standard errors are in parentheses.

Object	Estimate	Bootstrap SE	95% interval	Share positive
First-week avg. own sensitivity, no risk	5.038	0.295	[4.385, 5.565]	1.000
First-week avg. own sensitivity, with risk	5.049	0.296	[4.391, 5.576]	1.000
Paired first-week risk adjustment	0.011	0.002	[0.006, 0.015]	1.000
Weeks 2–4 systematic return sensitivity	8.167	1.330	[4.801, 10.010]	1.000
Weeks 2–4 avg. own-variance loading	0.188	0.286	[0.000, 1.018]	0.970

Table A.11. Bootstrap inference for dynamic demand
The table reports functionals of the 84-parameter with-risk system and its 64-parameter no-risk counterpart; both use 442 moments. Intervals are 95% percentile intervals from 200 twelve-week block-bootstrap draws. "Share positive" is the fraction of draws strictly above the zero boundary. The risk adjustment is calculated draw by draw as the with-risk minus no-risk first-week own sensitivity.

B Recovery with Institutional Holdings

The same method can recover investor-specific kernels when holdings responses are observed. Let investor i have demand

$$q_{i,t} = -A_{0,i} p_t + \sum_{s=1}^{S} A_{s,i} \mu^p_{t,t+s} + \sum_{s=1}^{S} B_{s,i} \mu^d_{t,t+s} - \sum_{s=1}^{S} C_{s,i} v^r_{t,t+s} + \xi_{i,t}.$$

Let $\Psi^{q_i}_h$ be the response of investor i's log holdings to the identified innovation. If $\xi_{i,t}$ satisfies the same exclusion restriction as aggregate latent demand, then

$$A_{0,i} \Psi^p_h - \sum_{s=1}^{S} A_{s,i} \Psi^p_{h+s} = \sum_{s=1}^{S} B_{s,i} \Phi^d_{s,h} - \sum_{s=1}^{S} C_{s,i} \Omega_{s,h} - \Psi^{q_i}_h. \tag{A.3}$$

The aggregate return, dividend, and risk responses are unchanged; the additional input is the investor's holdings-response path. Stacking (A.3) across horizons therefore recovers the investor-specific kernel under the same order and rank conditions.

This investor-level system aggregates back to the representative-investor equation. If $\bar{Q}_i$ is investor i's reference holding and $\bar{Q} = \sum_i \bar{Q}_i$, linearized market clearing implies

$$\sum_i \mathrm{Diag}(\bar{Q}_i) \Psi^{q_i}_h = -\mathrm{Diag}(\bar{Q}) \Psi^z_h.$$

Premultiplying each investor equation by $\mathrm{Diag}(\bar{Q}_i)$, summing, and dividing by $\mathrm{Diag}(\bar{Q})$ yields (A.4), with each aggregate structural matrix equal to its holdings-weighted investor average.

C Characteristic Definitions

This appendix lists the 120 monthly stock characteristics used to construct the five-factor basis. We use 106 characteristics directly from the Jensen-Kelly-Pedersen global factor data and 14 additional variables constructed from Jensen-Kelly-Pedersen inputs so that these transformed variables are not themselves price-scaled characteristics. All underlying Jensen-Kelly-Pedersen variable names and definitions follow Jensen et al. (2023) and the accompanying Global Factor Data Documentation.

Variable	Description
Jensen-Kelly-Pedersen characteristics used directly	
cowc_gr1a	Change in current operating working capital
oaccruals_at	Operating accruals scaled by assets
oaccruals_ni	Operating accruals scaled by net income
taccruals_at	Total accruals scaled by assets
taccruals_ni	Total accruals scaled by net income
capex_abn	Abnormal corporate investment
debt_gr3	Total debt growth over three years
fnl_gr1a	Change in financial liabilities over one year
ncol_gr1a	Change in non-current operating liabilities over one year
nfna_gr1a	Change in net financial assets over one year
ni_ar1	Lagged net income to assets
noa_at	Net operating assets scaled by total assets
aliq_at	Liquidity scaled by lagged assets
at_gr1	Asset growth over one year
be_gr1a	Change in book equity scaled by assets
capx_gr1	Capital expenditure growth over one year
capx_gr2	Capital expenditure growth over two years
capx_gr3	Capital expenditure growth over three years
coa_gr1a	Change in current operating assets over one year
col_gr1a	Change in current operating liabilities over one year
emp_gr1	Employee growth over one year
inv_gr1	Inventory growth over one year
inv_gr1a	Change in inventory over one year
lnoa_gr1a	Change in long-term net operating assets
mispricing_mgmt	Management-based mispricing
ncoa_gr1a	Change in non-current operating assets over one year
nncoa_gr1a	Change in net non-current operating assets over one year
noa_gr1a	Change in net operating assets over one year
ppeinv_gr1a	Change in property, plant, equipment, and inventories
sale_gr1	Sales growth over one year
sale_gr3	Sales growth over three years
saleq_gr1	Quarterly sales growth

Table A.12. Stock characteristics used in the five-factor recovery basis

Variable	Description
Jensen-Kelly-Pedersen characteristics used directly, continued	
age	Firm age
at_be	Book leverage
bidaskhl_21d	21-day bid-ask high-low spread
cash_at	Cash scaled by assets
ni_ivol	Earnings volatility
rd_sale	Research and development scaled by sales
rd5_at	Research and development capital scaled by assets
tangibility	Asset tangibility
beta_60m	60-month CAPM beta
beta_dimson_21d	21-day Dimson beta
betabab_1260d	Betting-against-beta
betadown_252d	252-day downside beta
earnings_variability	Earnings variability
ivol_capm_21d	21-day CAPM idiosyncratic volatility
ivol_capm_252d	252-day CAPM idiosyncratic volatility
ivol_ff3_21d	21-day Fama-French three-factor idiosyncratic volatility
ivol_hxz4_21d	21-day HXZ four-factor idiosyncratic volatility
ocfq_saleq_std	Volatility of quarterly operating cash flow to sales
rvol_21d	21-day return volatility
turnover_126d	126-day share turnover
zero_trades_21d	21-day zero-trade measure
zero_trades_126d	126-day zero-trade measure
zero_trades_252d	252-day zero-trade measure
dsale_dinv	Change in sales minus change in inventory
dsale_drec	Change in sales minus change in receivables
dsale_dsga	Change in sales minus change in SG&A
niq_at_chg1	Change in quarterly net income scaled by assets
niq_be_chg1	Change in quarterly net income scaled by book equity
niq_su	Quarterly earnings surprise
ocf_at_chg1	Change in operating cash flow scaled by assets
sale_emp_gr1	Sales growth relative to employee growth
saleq_su	Quarterly revenue surprise
tax_gr1a	Change in effective tax rate over one year

Table A.13. Stock characteristics used in the five-factor recovery basis, continued

Variable	Description
Jensen-Kelly-Pedersen characteristics used directly, continued	
dolvol_var_126d	126-day dollar-volume volatility
ebit_bev	EBIT scaled by business enterprise value
ebit_sale	EBIT scaled by sales
f_score	Piotroski F-score
ni_be	Net income scaled by book equity
niq_be	Quarterly net income scaled by book equity
o_score	O-score
ocf_at	Operating cash flow scaled by assets
ope_be	Operating profitability scaled by book equity
ope_bel1	Operating profitability scaled by lagged book equity
turnover_var_126d	126-day turnover volatility
at_turnover	Asset turnover
cop_at	Cash-based operating profitability scaled by assets
cop_atl1	Cash-based operating profitability scaled by lagged assets
dgp_dsale	Change in gross profit minus change in sales
gp_at	Gross profits scaled by assets
gp_atl1	Gross profits scaled by lagged assets
ni_inc8q	Number of consecutive earnings increases
niq_at	Quarterly net income scaled by assets
op_at	Operating profits scaled by assets
op_atl1	Operating profits scaled by lagged assets
opex_at	Operating leverage

Table A.14. Stock characteristics used in the five-factor recovery basis, continued

Variable	Description
Jensen-Kelly-Pedersen characteristics used directly, continued	
qmj	Quality minus junk
qmj_growth	Quality minus junk: growth
qmj_prof	Quality minus junk: profitability
qmj_safety	Quality minus junk: safety
sale_bev	Sales scaled by business enterprise value
corr_1260d	Market correlation
coskew_21d	Coskewness
dbnetis_at	Net debt issuance scaled by assets
lti_gr1a	Change in long-term investments over one year
pi_nix	Taxable income scaled by net income
sti_gr1a	Change in short-term investments over one year
ami_126d	Amihud illiquidity measure
iskew_capm_21d	21-day CAPM idiosyncratic skewness
iskew_ff3_21d	21-day Fama-French three-factor idiosyncratic skewness
iskew_hxz4_21d	21-day HXZ four-factor idiosyncratic skewness
rskew_21d	21-day return skewness
chcsho_12m	12-month change in shares outstanding
eqnetis_at	Equity net issuance scaled by assets
netis_at	Net issuance scaled by assets

Table A.15. Stock characteristics used in the five-factor recovery basis, continued

Variable	Description
Other variables (not as market-equity-scaled ratios as in Jensen et al., 2023)	
`netdebt`	Net debt
`rd`	Research and development expense
`at`	Total assets
`be`	Book equity
`debt`	Total debt
`div12m`	Trailing 12-month dividends
`ebitda`	EBITDA
`eqnpo`	Equity net payout
`eqpo`	Net equity payout
`fcf`	Free cash flow
`ival`	Intrinsic value
`ni`	Net income
`ocf`	Operating cash flow
`sale`	Sales

Table A.16. Additional accounting characteristics used in the five-factor recovery basis

Notes: The first three tables report the 106 characteristics taken directly from the Jensen et al. (2023) global factor data. The final table reports the 14 additional characteristics we construct by recovering underlying accounting quantities from Jensen-Kelly-Pedersen variables that are distributed in market-equity-scaled form. These transformations ensure that the variables in the final table enter the factor-model instrumentation as underlying accounting quantities rather than price-scaled ratios. Characteristics are lagged by one month when matched to returns, and each month we apply the rank-to-percentile transform and recenter at zero.

D Proofs

D.1 Proof of Lemma 2

Proof. Let F denote the forward-shift operator on the projection coefficients, so $(F\Psi)_h = \Psi_{h+1}$. The dynamic recovery equations can then be written as

$$\Psi^z = \mathcal{A}(F)\Psi^p, \qquad \mathcal{A}(F) \equiv A_0 - \sum_{s=1}^{S} A_s F^s.$$

This identity implies that every row of $\mathcal{H}(\Psi^z)$ is a finite linear combination of rows of $\mathcal{H}(\Psi^p)$. Hence, $\operatorname{rank}\mathcal{H}(\Psi^z) \le \operatorname{rank}\mathcal{H}(\Psi^p)$. Under Assumption 2, Proposition 1 gives $\Psi^p = \mathcal{A}(F)^{-1}\Psi^z$, where the inverse is absolutely convergent. Every row of $\mathcal{H}(\Psi^p)$ is therefore an absolutely convergent linear combination of rows of $\mathcal{H}(\Psi^z)$. Because the row space of $\mathcal{H}(\Psi^z)$ is finite-dimensional and hence closed, $\operatorname{rank}\mathcal{H}(\Psi^p) \le \operatorname{rank}\mathcal{H}(\Psi^z)$. Combining the two inequalities gives the result. □

D.2 Proof of Corollary 1

Proof. Under stable invertibility, the price response associated with (20) is

$$\Psi^p_h = \sum_{k=1}^{K} b_k \rho_k^h, \qquad b_k \equiv \frac{c_k}{\mathcal{A}(\rho_k)}.$$

Let V_m denote the $(m+1)\times K$ Vandermonde matrix with (j,k) entry ρ_k^{j-1}. The recovery matrix factors as

$$\overline{\Psi}^p = V_S \operatorname{Diag}(b_1,\dots,b_K) V_H^\top.$$

Because the roots are distinct and the loadings are nonzero, the two Vandermonde matrices have ranks $\min\{S+1,K\}$ and $\min\{H+1,K\}$, while the diagonal matrix is nonsingular. Because $H \ge K-1$, V_H has full column rank K. Postmultiplication by $V_H^\top$ therefore preserves the rank of $V_S \operatorname{Diag}(b_1,\dots,b_K)$. Since V_S has rank $\min\{S+1,K\}$ and the diagonal matrix is nonsingular, $\operatorname{rank}\overline{\Psi}^p = \min\{K,S+1\}$. □

E Recovery of the Full Dynamic Demand Kernel

We now restore the expected-dividend and risk terms in (1). Let Ψ^d_h denote the coefficient from projecting d_t on ι_{t-h}, and let $\Psi^v_{s,h}$ denote the corresponding coefficient for the horizon-s risk term $v^r_{t,t+s}$. By iterated expectations, the coefficient from projecting $\mathbb{E}_t[d_{t+s}]$ on ι_{t-h} is Ψ^d_{h+s}. Projecting the full equilibrium condition therefore gives

$$A_0\Psi^p_h - \sum_{s=1}^{S} A_s\Psi^p_{h+s} - \sum_{s=1}^{S} B_s\Psi^d_{h+s} + \sum_{s=1}^{S} C_s\Psi^v_{s,h} = \Psi^z_h, \qquad h \geq 0. \tag{A.4}$$

Define

$$\Theta \equiv \begin{bmatrix} A_0 & -A_1 & \cdots & -A_S & -B_1 & \cdots & -B_S & C_1 & \cdots & C_S \end{bmatrix}$$

and

$$X_h \equiv \begin{bmatrix} (\Psi^p_h)^\top & (\Psi^p_{h+1})^\top & \cdots & (\Psi^p_{h+S})^\top & (\Psi^d_{h+1})^\top & \cdots & (\Psi^d_{h+S})^\top & (\Psi^v_{1,h})^\top & \cdots & (\Psi^v_{S,h})^\top \end{bmatrix}^\top.$$

Stacking X_h across horizons gives $X_H \equiv [X_0, \ldots, X_H]$ and $Z_H \equiv [\Psi^z_0, \ldots, \Psi^z_H]$.

Proposition 4 (Recovery of the full dynamic demand kernel)**.** *Suppose Assumption 1 holds and the population response coefficients entering* (A.4) *are known. If X_H has full row rank, then the dynamic demand kernel is identified and*

$$\Theta = Z_H X_H^\top (X_H X_H^\top)^{-1}.$$

The proposition permits each instrument to move multiple prices, dividends, risk components, assets, and horizons simultaneously. What matters is whether the resulting paths jointly span the arguments needed to recover the kernel. The econometrician does not need a separate instrument for each expected variable, does not need to observe investors' expectations, and does not need to recover the full state or structural-shock vector from an invertible state-space system.

F Additional Dynamic-Rank Examples

For a coprime ARMA(p,q) response, the autoregressive component generates p persistent propagation patterns, while the moving-average component can contribute independent variation

at the initial horizons. A generic ARMA(p,q) response has Hankel rank $\max\{p,q+1\}$. Once the moving-average component has left the selected horizon block, however, the rank of that block is at most p. This also clarifies why invertibility of a moving-average polynomial does not create additional identifying variation. An invertible MA(q) has an AR(∞) representation, but its response still has Hankel rank at most $q+1$. Identification depends on the response sequence rather than the nominal length of an observationally equivalent autoregressive representation.

Figure A.1 illustrates the role of horizon selection using an ARMA(1,2) response. The recovery matrix formed from $h = 0,1,2$ retains the moving-average component and has rank three. Starting the three-equation block at $h = 1$ reduces its rank to two, and starting at $h = 2$ or later leaves only the single persistent autoregressive component. Thus, the same instrument-induced response can contain different amounts of identifying variation at different portions of its path.

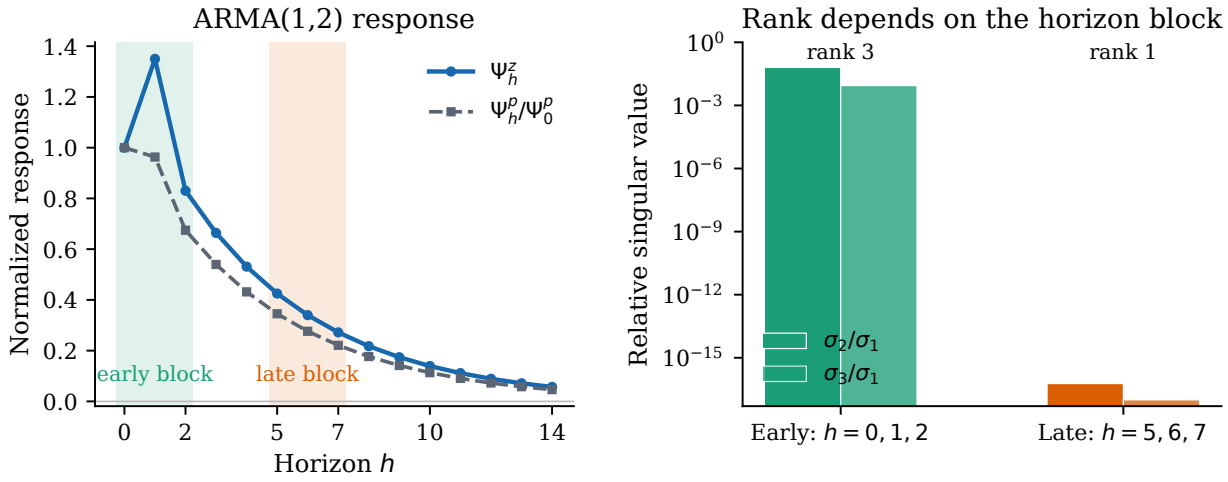


Figure A.1. The location of identifying variation in an ARMA response

Notes: The figure considers an ARMA(1,2) outside-demand response with autoregressive coefficient 0.8 and moving-average coefficients (1,0.55, – 0.25). The structural demand coefficients are $(A_0,A_1,A_2) = (5,2.5,1)$. The left panel shows the outside-demand and normalized equilibrium-price responses and highlights early and late three-equation blocks. The right panel reports the normalized singular values of the corresponding recovery matrices.

G Integrability and Demand Coefficients

The integrability theorem (Proposition 3.H.1 in Mas-Colell, Whinston, and Green, 1995) states that a continuously differentiable demand function can be rationalized by a continuous, locally

nonsatiated, and strictly quasiconcave utility function if and only if: (i) the budget constraint holds with equality (Walras' law), (ii) demand is homogeneous of degree zero in prices and wealth, and (iii) the Slutsky substitution matrix is symmetric and negative semi-definite. The symmetry of the Slutsky matrix is the economically substantive condition. It requires that the compensated cross-price effects be equal: the compensated effect on the demand for asset i of a change in the price of asset j must equal the compensated effect on asset j of a change in the price of asset i.

In our setting, the demand arguments are current prices p_t and the sequences of conditional expectations $(\mu^p_{t,t+s}, \mu^d_{t,t+s}, v^r_{t,t+s})$. The Jacobian of q_t with respect to these arguments is composed of the coefficient matrices: $-A_0$ is the own-price derivative, A_s is the derivative with respect to $\mu^p_{t,t+s}$, B_s is the derivative with respect to $\mu^d_{t,t+s}$, and $-C_s$ is the derivative with respect to $v^r_{t,t+s}$. The integrability conditions therefore impose cross-equation restrictions across these matrices. For instance, if demand is generated by myopic mean-variance preferences with a symmetric covariance matrix Σ, then A_0 inherits the symmetry and positive definiteness of Σ^{-1}. More generally, symmetry of the compensated substitution effects constrains the relationship between the own-price matrix A_0 and the intertemporal price coefficients A_s.

Example: CARA/mean–variance demand. To make the integrability conditions concrete, consider the myopic CARA investor from Section 2.3. The investor maximizes $\mathbb{E}_t[-\exp(-\gamma\, W_{t+1})]$ subject to $W_{t+1} = R^f(W_t - P_t^\top Q_t) + (P_{t+1} + \mathcal{D}_{t+1})^\top Q_t$, where $\gamma > 0$ is the coefficient of absolute risk aversion. Under conditional normality of returns with covariance $\Sigma \succ 0$, the first-order condition yields

$$Q_t = \frac{1}{\gamma}\,\Sigma^{-1}\big(\mathbb{E}_t[P_{t+1} + \mathcal{D}_{t+1}] - R^f\, P_t\big).$$

We verify the integrability conditions for this demand and trace the restrictions they impose on the coefficient matrices.

Slutsky symmetry and negative semi-definiteness. Under CARA preferences demand does not depend on current wealth W_t, so the income effect vanishes: $\partial Q_t / \partial W_t = 0$. The Slutsky substitution matrix therefore coincides with the Marshallian (uncompensated) price derivative:

$$S \;\equiv\; \frac{\partial Q_t}{\partial P_t^\top} + \frac{\partial Q_t}{\partial W_t}\, Q_t^\top \;=\; \frac{\partial Q_t}{\partial P_t^\top} \;=\; -\frac{R^f}{\gamma}\,\Sigma^{-1}.$$

Because Σ is symmetric and positive definite, so is Σ^{-1}. Therefore $S = -(R^f/\gamma)\,\Sigma^{-1}$ is

(i) symmetric and (ii) negative definite, verifying the two economically substantive integrability conditions. Symmetry says that the compensated effect on demand for asset i of a change in the price of asset j equals the compensated effect on asset j of a change in the price of asset i. This is a testable restriction whenever $N \geq 2$.

Cross-equation restrictions on the log-linear coefficients. Log-linearizing the demand around steady state $(\bar{Q}, \bar{P}, \bar{\mathcal{D}})$ yields the coefficient matrices:

$$A_0 = \mathrm{Diag}(\bar{Q})^{-1} \frac{R^f}{\gamma} \Sigma^{-1} \mathrm{Diag}(\bar{P}), \qquad A_1 = \beta A_0, \qquad B_1 = \mathrm{Diag}(\bar{Q})^{-1} \frac{1}{\gamma} \Sigma^{-1} \mathrm{Diag}(\bar{\mathcal{D}}),$$

where $\beta = 1/R^f$. All higher-order coefficients vanish: $A_s = B_s = C_s = 0$ for $s \geq 2$. These coefficients are not free: the CARA utility imposes several cross-equation restrictions.

First, the relation $A_1 = \beta A_0$ ties the current-price and expected-future-price coefficients together. Because the investor cares about prices only through one-period expected excess payoffs $\mathbb{E}_t[P_{t+1} + \mathcal{D}_{t+1}] - R^f P_t$, a unit increase in p_t with an offsetting R^f-unit increase in $\mathbb{E}_t[p_{t+1}]$ has no effect on demand.

Second, the rescaled Jacobian $\tilde{A}_0 \equiv \mathrm{Diag}(\bar{Q}) A_0 \mathrm{Diag}(\bar{P})^{-1} = (R^f/\gamma)\, \Sigma^{-1}$ must be symmetric. This is the Slutsky symmetry condition expressed in the log-linear coefficients, and it holds because Σ^{-1} is symmetric. When A_0 is estimated freely with N^2 parameters, the $N(N-1)/2$ symmetry conditions $(\tilde{A}_0)_{ij} = (\tilde{A}_0)_{ji}$ for $i \neq j$ are overidentifying restrictions that can be tested empirically.

Third, the vanishing of all coefficients beyond $s = 1$ is itself a restriction. A myopic investor has no hedging motive and faces no adjustment costs, so expectations beyond the next period do not enter demand.

Recovering preferences from demand coefficients. The integrability theorem says not only that the conditions above are necessary, but that they are sufficient to recover the generating preferences. For the CARA case the recovery is explicit. Given A_0 and the observables $(\bar{Q}, \bar{P}, R^f)$, one obtains

$$\frac{1}{\gamma} \Sigma^{-1} \;=\; \frac{1}{R^f} \mathrm{Diag}(\bar{Q}) A_0 \mathrm{Diag}(\bar{P})^{-1},$$

which identifies the composite object $\gamma^{-1}\Sigma^{-1}$. If the return covariance Σ is separately observable from return data, inverting gives the risk-aversion parameter γ directly. One may also use B_1 to form a second expression, $\gamma^{-1}\Sigma^{-1} = \mathrm{Diag}(\bar{Q}) B_1 \mathrm{Diag}(\bar{\mathcal{D}})^{-1}$, and test whether the two routes to $\gamma^{-1}\Sigma^{-1}$ agree. Rejection would indicate a departure from the CARA specification.

The example illustrates how integrability turns observed demand sensitivities into statements about preferences. Symmetry of the Slutsky matrix disciplines the shape of the investor's risk/return trade-off. The restriction $A_1 = \beta A_0$ pins down the discount rate. And the fact that all higher-order coefficients vanish tells us the investor's planning horizon is one period.

H Sequential and Recursive Equilibrium Representations

This appendix expands on the discussion in Section 2.2 regarding the sequential and recursive representations of equilibria and the sense in which the sequential formulation is more general.

Two representations of complete-market equilibria. A classical result in dynamic economic theory is that, under complete markets, competitive equilibria can be characterized in two ways (Ljungqvist and Sargent, 2018, Chapter 8). In the Arrow–Debreu formulation, all agents meet at date 0 and trade a complete set of state-contingent claims, one for every date-event pair (t,s^t). Each agent solves a single lifetime problem subject to a single budget constraint. In the *sequential* formulation, agents trade period by period in spot markets, choosing portfolios of one-period Arrow securities at each date t and history s^t. Each agent solves a dynamic programming problem, and equilibrium requires market clearing at every date-event pair.

Under complete markets the two formulations deliver identical allocations and prices. The Arrow–Debreu formulation is conceptually clean but requires specifying the full probability tree up front; the sequential formulation is more natural for dynamic models and connects directly to recursive methods.

The sequential representation in detail. The demand system (1) is written in sequential form. At each date t, the investor conditions on the filtration generated by v^t and forms expectations about the entire future path of prices $\{\mu^p_{t,t+s}\}_{s=1}^S$, dividends $\{\mu^d_{t,t+s}\}_{s=1}^S$, and return risk $\{v^r_{t,t+s}\}_{s=1}^S$. The demand system is therefore a map from the full sequence of conditional expectations to current holdings.

This representation has two important features. First, the entire sequence of expected prices, dividends, and volatilities shows up explicitly as arguments of the demand function. There is no assumption that these sequences are generated by a low-dimensional state vector; they are treated as free inputs. Second, combining the demand system with market clearing directly links the

structural demand coefficients $\{A_s\}$ to the impulse responses of equilibrium prices through the recovery equation (A.4). This equation must hold at every horizon h and provides the basis for recovering the structural demand matrices from observed impulse responses.

The recursive competitive equilibrium alternative. An alternative approach, when applicable, is to cast the equilibrium as a *recursive competitive equilibrium* (RCE). In the RCE formulation, one assumes that the relevant state of the economy at date t can be summarized by a finite-dimensional Markov state vector s_t. Value functions and policy functions are then expressed as functions of s_t alone, and the equilibrium is characterized as a fixed point in the space of policy and pricing functions.

As an example of the recursive approach applied to the study of demand elasticities in dynamic settings, Binsbergen et al. (2025) study a continuous-time economy with CRRA investors and a risk-free asset plus J risky assets. Absent exogenous price-shifter shocks, asset prices follow diffusions with aggregate and idiosyncratic Brownian components. A Markov "shifter state" $s_t \in \Omega$ drives exogenous log price shifts $\beta_j s_t$ (interpreted as arising from noise traders, central banks, or share issuances), so the actual trading price of asset j is $P_{j,t} = e^{\beta_j s_t} \tilde{P}_{j,t}$. Events arrive at Poisson rate λ; at each event date the shifter state transitions according to $q(s'|s)$, agents may consume and be replaced, and portfolios are rebalanced. The value function takes the CRRA form $V(W,s) = u(W)\, A(s)$, and optimal portfolio shares $\theta^*(s)$ solve the HJB equation jointly with the function $A(s)$.

The key insight of Binsbergen et al. (2025) is that the shifter state s_t alters the entire future path of resale prices, so a shock that shifts the contemporaneous price while also moving future prices need not identify a structural price elasticity. They show that, under continuous trading, true own-price elasticities are infinite (Proposition 1 of their paper), while discrete trading frictions produce finite elasticities that are governed by the trading frequency and the resolution probability of the price shift. The estimated "shifter-process elasticity" is specific to the dynamics of the particular shock process and therefore does not represent a structural object.

The RCE approach has clear advantages: once the state vector and functional forms are specified, the model can be solved numerically and used for counterfactual analysis. However, it requires the researcher to commit to a particular set of state variables and to specify how they enter preferences and technology. The choice of state vector effectively determines the model.

Why the sequential representation. The sequential representation is strictly more general than the recursive alternative, and this generality is central to our approach.

First, the sequential demand system (1) does not require specifying or identifying the state variables that drive expectations. The structural coefficients $\{A_s, B_s, C_s\}$ characterize how demand responds to expectations at each horizon, regardless of which underlying states generate those expectations. Any RCE with a well-defined demand function is a special case of the sequential formulation, but the converse is not true: the sequential form accommodates settings where no finite Markov state suffices. This is what makes the structural elasticities transportable across shock environments. They are properties of the investor's demand function, not of a particular equilibrium model.

Second, the sequential representation connects directly to the impulse-response objects that are estimable from data. The convolution equation (A.4) links the structural demand matrices to the reduced-form impulse responses of prices and flows. Because the sequential form keeps the entire expected path as arguments, the dynamic structure of the impulse responses can be exploited for identification without assuming a particular law of motion for the state.

Third, nothing is lost by working at this level of generality. Any recursive model that satisfies the conditions for the Ljungqvist–Sargent equivalence admits an equivalent sequential representation in which the demand coefficients $\{A_s, B_s, C_s\}$ are well-defined. The structural elasticities recovered from the sequential representation are therefore the same objects that would emerge from a correctly specified recursive model.

I Numerical Illustration of Return-Based Recovery

This appendix gives a scalar version of the one-step benchmark calculation. The purpose is to make the sign convention and the role of the return response transparent. The empirical recovery in Section 6, including the preferred four-week specification, uses the full fixed-weight factor system, all included horizons, and both the trace and factor-block moments.

Suppress factor notation and average the stock-level diagonal restrictions. In the one-step specification without risk, the horizon-zero recovery equation is

$$0 = \bar{\Psi}_0^z + D_1 \bar{\Psi}_1^r, \qquad D_1 = -\frac{\bar{\Psi}_0^z}{\bar{\Psi}_1^r}.$$

Here $\bar{\Psi}_0^z$ is the average contemporaneous outside-demand response, and $\bar{\Psi}_1^r$ is the forward return from week $t+1$ associated with a shock in week t. It is not the cumulative return through week $t+1$. This distinction is central. The impact return is the contemporaneous price response to the shock; the one-step structural coefficient is identified from the next-period expected return that investors anticipate when choosing demand at date t.

The five-factor first-stage estimates in Table A.5 give

$$\bar{\Psi}_0^z = 1.051, \qquad \bar{\Psi}_1^r = -0.184.$$

The scalar calculation is therefore

$$D_1 \approx -\frac{1.051}{-0.184} = 5.71.$$

This is close to the full five-factor no-risk estimate of 5.167 in Table A.6. The two numbers need not be identical because the full estimate uses all moments in the fixed-weight factor system rather than only the single averaged diagonal equation. The same calculation in the one-factor basis uses $\bar{\Psi}_0^z = 1.053$ and $\bar{\Psi}_1^r = -0.169$, giving

$$-\frac{1.053}{-0.169} = 6.23,$$

again matching the scale of the full one-factor no-risk estimate of 5.520 in Table A.2.

With risk, the scalar one-step equation becomes

$$0 = \bar{\Psi}_0^z + D_1 \bar{\Psi}_1^r - \overline{C_1^F \Omega_1}, \qquad D_1 = \frac{\overline{C_1^F \Omega_1} - \bar{\Psi}_0^z}{\bar{\Psi}_1^r}. \tag{A.5}$$

In the five-factor estimates, the projected factor-risk response is negative on average and the row-PSD restriction makes the economically relevant risk loading positive. Thus $\overline{C_1^F \Omega_1}$ is negative, so the numerator in (A.5) is more negative than in the no-risk calculation. Since $\bar{\Psi}_1^r < 0$, including risk raises the recovered value of D_1. This is the basic algebra behind the one-step benchmark increase from 5.167 without risk to 5.188 with risk in Table A.6, and from 5.520 to 6.741 in Table A.2. The smaller risk adjustment in the preferred dynamic specification obeys the same sign logic while allowing the expected-return and own-variance kernels to vary separately across four weeks.

J Identification of Dynamic Causal Effects with External Instruments

This appendix states the population conditions under which the impulse-response inputs used in the recovery system are identified as dynamic causal effects of a demand shock. The notation follows Stock and Watson (2018) but is specialized to the demand-system setting of the paper.

J.1 Population objects

Let y_t collect the recovery outcomes of interest. In the unrestricted notation,

$$y_t \equiv \left(z_t^\top, p_t^\top, \{\mu_{t,t+s}^{d\,\top}\}_{s=1}^{S}, \{v_{t,t+s}^{r\,\top}\}_{s=1}^{S}\right)^\top.$$

In the return-based empirical implementation, p_t is replaced by forward excess returns and the risk state is the factor realized second moment. Suppose the recovery outcomes admit the structural moving-average representation

$$y_t = \bar{y} + \sum_{\ell=0}^{\infty} \Theta_\ell^{\varepsilon} \varepsilon_{t-\ell} + \sum_{\ell=0}^{\infty} \Theta_\ell^{\chi} \chi_{t-\ell}. \tag{A.6}$$

The scalar ε_t is the identified outside-demand innovation used throughout the recovery section, and χ_t collects all other structural shocks. The dynamic causal effect of a one-unit demand innovation on the recovery outcomes at horizon h is

$$\mathrm{IRF}_h^d \equiv \frac{\partial y_{t+h}}{\partial \varepsilon_t} = \Theta_h^{\varepsilon}. \tag{A.7}$$

The components of (A.7) are the population objects denoted by Ψ_h^z, Ψ_h^r, $\Phi_{s,h}^d$, and $\Omega_{s,h}$ in Section 5.

J.2 External-instrument conditions

Identification requires the following population restrictions on the identified innovation ε_t. In the panel implementation, the corresponding object is the stock-level innovation $\iota_{j,t}$ after residualizing with respect to the controls in the LP.

Assumption 3 (External-instrument validity)**.** *The identified demand innovation ε_t satisfies:*

(i) Relevance: $\Psi_0^z \neq 0$*, so the innovation moves outside demand on impact.*

(ii) Innovation normalization: $\mathbb{E}[\varepsilon_t] = 0$*,* $\mathbb{E}[\varepsilon_t^2] > 0$*, and* $\mathbb{E}[\varepsilon_t \varepsilon_{t+j}] = 0$ *for every* $j \neq 0$*.*

(iii) Contemporaneous exclusion: $\mathbb{E}[\varepsilon_t \chi_t^\top] = 0$*.*

(iv) Lead-lag exclusion: $\mathbb{E}[\varepsilon_t \chi_{t+j}^\top] = 0$ *for every* $j \neq 0$*.*

(v) Predetermined controls: *when the projection includes controls w_t, the preceding orthogonality conditions hold for the residualized innovation $\tilde{\varepsilon}_t = \varepsilon_t - \mathbb{E}[\varepsilon_t \mid w_t]$ and the residualized outcome.*

(vi) Demand-system exclusion: *the latent demand shifter has no projection on the instrumented demand-shock history, so $\Psi_h^\xi = 0$ for all $h \geq 0$ in Assumption 1.*

The first four restrictions are the standard external-instrument assumptions, written after normalizing the observed proxy into the identified innovation ε_t. Relevance says the innovation moves outside demand. Contemporaneous exclusion rules out correlation with non-demand structural shocks at date t. Lead-lag exclusion rules out anticipation and delayed correlation with non-demand shocks at other dates. The fifth condition is the controlled version used in empirical LPs and VARs. The final condition is specific to the demand-system recovery: ε_t may move equilibrium prices, returns, and risk through the identified demand innovation, but it cannot directly move the unobserved elastic-investor demand shifter ξ_t.

J.3 IV LPs

For a scalar outcome component $y_{i,t+h}$, define the population LP coefficient

$$\beta_{i,h}^{LP} \equiv \frac{\mathbb{E}[y_{i,t+h}\tilde{\varepsilon}_t]}{\mathbb{E}[\tilde{\varepsilon}_t^2]},$$

where $\tilde{\varepsilon}_t = \varepsilon_t$ if no predetermined controls are included. Substituting the moving-average representation (A.6) gives

$$\mathbb{E}[y_{i,t+h}\tilde{\varepsilon}_t] = \sum_{\ell=0}^{\infty} e_i^\top \Theta_\ell^\varepsilon \mathbb{E}[\varepsilon_{t+h-\ell}\tilde{\varepsilon}_t] + \sum_{\ell=0}^{\infty} e_i^\top \Theta_\ell^\chi \mathbb{E}[\chi_{t+h-\ell}\tilde{\varepsilon}_t].$$

By Assumption 3, every term is zero except the demand-innovation term with $\ell = h$. Therefore

$$\beta_{i,h}^{LP} = e_i^\top \Theta_h^\varepsilon.$$

Thus LP-IV identifies the structural impulse response at each horizon in the units of the identified innovation ε_t. If a raw external proxy is used instead, its normalization only rescales all impulse responses by a common constant. The recovery equation is invariant to this common shock scaling because the flow, return, dividend, and risk responses all refer to the same normalized innovation.

In practice, the LP also includes lagged outcomes or other predetermined variables. These controls are not part of the causal object; they residualize the innovation and outcome so that Assumption 3 holds conditionally. Under the same population restrictions, adding sufficient lag controls leaves the target impulse response unchanged, as in Stock and Watson (2018) and Plagborg-Møller and Wolf (2021).

J.4 Proxy structural VAR

Alternatively, suppose the recovery outcomes are represented by a finite-order reduced-form VAR,

$$y_t = \sum_{p=1}^{P} A_p y_{t-p} + u_t, \qquad u_t = b_\varepsilon \varepsilon_t + B_\chi \chi_t,$$

where b_ε is the impact effect of the identified demand innovation on the VAR innovations. Under Assumption 3,

$$\mathbb{E}[\varepsilon_t u_t] = \mathbb{E}[\varepsilon_t^2]\, b_\varepsilon.$$

Hence the covariance between the identified innovation and the reduced-form VAR residual identifies the demand-shock impact vector after dividing by $\mathbb{E}[\varepsilon_t^2]$. Iterating the VAR then gives

$$\mathrm{IRF}_h^{VAR} = J \mathcal{A}^h J^\top b_\varepsilon,$$

where $\mathcal{A}$ is the companion matrix and J selects the contemporaneous block. This is the proxy-SVAR estimator used in the external-instrument literature.

Plagborg-Møller and Wolf (2021) show that, with unrestricted lag structures, VARs and LPs estimate the same population impulse responses for any identification scheme, including external

instruments. With a finite lag order P, a correctly specified VAR can be more efficient because it imposes a law of motion, while LPs estimate each horizon directly and are less sensitive to dynamic misspecification. This tradeoff affects the precision and robustness of the impulse-response inputs, not the definition of the recovery equation.

J.5 Panel implementation

The empirical design applies the same external-instrument logic in an unbalanced stock panel. For an outcome $y_{j,t+h}$ equal to outside demand or returns, the population version of the stock-level LP is

$$y_{j,t+h} = \alpha^{y}_{h,t} + \phi^{y\top}_{h}\ell_{j,t} + a^{(h)}_{y}\iota_{j,t} + \ell^{\top}_{j,t}G^{(h)}_{y}s_t + u^{y}_{j,t+h}, \qquad s_t = L^{\top}_{t}\iota_t. \tag{A.8}$$

This specification imposes that the stock-level response matrix lies in the factor-plus-own class

$$\mathrm{IRF}^{y}_{h}(L_t) = a^{(h)}_{y}I_{N_t} + L_tG^{(h)}_{y}L^{\top}_{t}.$$

The date fixed effect $\alpha^{y}_{h,t}$ absorbs aggregate shocks common to all stocks in week t, the loading control removes systematic differences associated with predetermined factor exposures, the own coefficient identifies the direct stock-level demand-shock effect, and $G^{(h)}_{y}$ identifies factor-mediated spillovers. The factor-risk projection (27) is the corresponding factor-level LP applied to the realized second-moment outcome after stock-level demand innovations have been aggregated to s_t.

The panel analogue of Assumption 3 requires the residualized stock-level innovation $\iota_{j,t}$ to be relevant for stock j's outside-demand shock and orthogonal, after week fixed effects and predetermined loading controls, to non-demand shocks affecting all recovery outcomes at all leads and lags. It also requires that the factor-plus-own response class is correctly specified for the dynamic causal effects used in recovery. If the true response matrix has components outside this class, the LPs identify the projection of the dynamic causal effects onto the factor-plus-own space, and the second-stage minimum-distance estimator recovers the demand coefficients for that projected system.

Finally, identification of the demand coefficients requires more than identification of the dynamic causal effects. The external-instrument assumptions identify the first-stage impulse-response inputs. The second-stage recovery additionally requires the rank condition in Proposition 4, or its factor-model analogue in Section 5: the estimated paths of flow, return, dividend, and risk responses

must contain enough independent horizon and cross-sectional variation to distinguish the primitive demand coefficients.

K Lagged Quantity and Partial Adjustment Robustness

This appendix addresses the possibility that demand depends directly on past quantity. This concern is distinct from forward-looking demand for expected returns and risk. A partial-adjustment investor carries past positions into the current portfolio because changing positions is costly; Van der Beck (2025), for example, uses lagged institutional holdings to distinguish short- from long-run demand elasticities. If the true demand equation contains lagged quantity but the econometrician omits it, the omitted state enters measured latent demand. Because a demand shock changes quantity and therefore future lagged quantity, this omission can violate Assumption 1 at horizons after impact.

K.1 Augmented recovery equation

We add one common partial-adjustment coefficient to the preferred return-based demand system:

$$q_t = \rho q_{t-1} + \sum_{s=1}^{4} D_{t,s}\mathbb{E}_t[r_{t+s}] - \sum_{s=1}^{4} L_t C_s^F \sigma_{t,t+s}^F + \xi_t, \qquad 0 \le \rho < 1. \tag{A.9}$$

The scalar restriction is deliberately parsimonious: it allows quantities to adjust gradually without introducing another matrix whose dimension grows with the number of stocks. We cap ρ at 0.95 in estimation. The expected-return matrices retain the restrictions in (24), and the risk coefficients retain the own-variance restriction in (25). We do not restrict the four weekly D_s^F matrices or the four vectors of risk coefficients to be proportional or monotone.

Let $\Psi_h^{q,F}$ denote the factor-space response of the elastic investor's quantity to the OFI innovation. Taking impulse responses of (A.9) gives

$$\Psi_h^{q,F} = \rho\Psi_{h-1}^{q,F} + \sum_{s=1}^{4} D_s^F \widehat{\Psi}_{h+s}^{r,F} - \sum_{s=1}^{4} C_s^F \widehat{\Omega}_{h+s}, \qquad \Psi_{-1}^{q,F} = 0.$$

Market clearing implies $\Psi_h^{q,F} = -\Psi_h^{z,F}$. The augmented factor recovery equation is therefore

$$-\left(\widehat{\Psi}_h^{z,F} - \rho\widehat{\Psi}_{h-1}^{z,F}\right) = \sum_{s=1}^{4} D_s^F \widehat{\Psi}_{h+s}^{r,F} - \sum_{s=1}^{4} C_s^F \widehat{\Omega}_{h+s}, \qquad h = 0, \ldots, 16, \tag{A.10}$$

where $\widehat{\Psi}_{-1}^{z,F} = 0$. Thus partial adjustment quasi-differences the quantity-side impulse response. Importantly, (A.10) introduces no new instrument. The same OFI innovation identifies $\widehat{\Psi}_h^{z,F}$ and $\widehat{\Psi}_{h-1}^{z,F}$, and ρ is identified from how their horizon profiles line up with the separately estimated return and risk response paths. This differs from inserting raw lagged quantity as an endogenous regressor in a levels regression.

We apply the same transformation to the stock-level factor-plus-own response. In the notation of (28), the augmented residual is

$$\Delta_h^{\rho}(\theta; L_t) \equiv \widehat{\Psi}_{t,h}^{z} - \rho\widehat{\Psi}_{t,h-1}^{z} + \sum_{s=1}^{4} D_{t,s}(\theta)\widehat{\Psi}_{t,h+s}^{r} - \sum_{s=1}^{4} L_t C_s^F(\theta)\widehat{\Omega}_{h+s} L_t'.$$

The projected moments are $N_t^{-1}\,\mathrm{tr}(\Delta_h^{\rho})$ and $(5/N_t)L_t'\Delta_h^{\rho}L_t$, exactly as in (29). All parameters enter linearly before the cone restrictions. Consequently, adding ρ preserves the convex quadratic minimum-distance problem: the no-risk system has 65 parameters, the risk system has 85, and both use 442 moments. We estimate all structural coefficients and ρ jointly.

The main specification treats the OFI response as the response of the outside-demand state z_t, as in the main recovery specification in the text. We also consider a deliberately conservative inventory mapping. If measured OFI were instead a flow that must be accumulated into an outside position, define

$$K_h^z \equiv \sum_{j=0}^{h} \Psi_j^z \tag{A.11}$$

and replace Ψ_h^z by K_h^z in (A.10). This alternative is useful as an accounting check, but it creates a mechanical force toward $\rho = 1$: $K_h^z - K_{h-1}^z = \Psi_h^z$. A near-unit estimate under this mapping should therefore not be read as preference inertia.

K.2 Predetermined-state controls and inference

As a second check, we allow the first-stage LPs to condition directly on pre-shock trading history. For both the OFI and return LPs, we augment (A.8) with

$$W_{n,t} = \left(\mathrm{OFI}_{n,t-1}, \ldots, \mathrm{OFI}_{n,t-4}, r_{n,t-1}, \ldots, r_{n,t-4}\right)'.$$

These eight variables enter as controls, not instruments; OFI remains the only identifying innovation. The factor-risk LP is unchanged because it is estimated from the factor-level weekly series. Requiring complete lag histories removes 70,978 observations from the 7,402,271-row stock panel. The largest absolute sample correlation between the residualized OFI innovation and any added lag is 0.00063.

Inference uses 200 twelve-week block-bootstrap draws. Every draw re-estimates the OFI, return, and risk impulse responses and then resolves the joint cone-constrained recovery problem. The PCA basis is held fixed, as in the main bootstrap. We report percentile intervals for boundary parameters because a Gaussian approximation is inappropriate when ρ, an eigenvalue, or a risk loading binds at zero.

	Level state	Predetermined lags	Cumulative inventory
$\widehat{\rho}$, no risk	0.000	0.000	0.950
	[0.000, 0.000]	[0.000, 0.000]	[0.950, 0.950]
$\widehat{\rho}$, with risk	0.000	0.000	0.950
	[0.000, 0.000]	[0.000, 0.000]	[0.950, 0.950]
First-week avg. own sensitivity	5.049	5.102	4.958
	[4.488, 5.605]	[4.465, 5.657]	[4.262, 5.409]
Paired first-week risk adjustment	0.011	0.011	0.002
	[0.006, 0.015]	[0.006, 0.015]	[−0.003, 0.005]
Weeks 2–4 systematic sensitivity	8.167	6.899	5.546
	[4.592, 10.727]	[4.147, 9.575]	[3.354, 8.122]
Weeks 2–4 avg. own-variance loading	0.188	0.154	0.153
	[0.000, 0.904]	[0.000, 0.981]	[0.000, 0.490]
Jacobian rank / parameters, with risk	85/85	85/85	85/85
Scaled condition number	260.4	252.7	262.7
Lagged-quantity VIF	1.65	1.60	5.42

Table A.17. Lagged-quantity robustness
This table summarizes the partial-adjustment estimates. The first column uses the maintained level-state interpretation of OFI; the second additionally includes four predetermined OFI and four predetermined return lags in the stock-level first stages; the third cumulates OFI responses into an inventory before estimating the recovery system. The demand parameters are from the with-risk specification unless otherwise indicated. Brackets report 95% percentile intervals from 200 twelve-week block-bootstrap draws. "Weeks 2–4 systematic sensitivity" is $\sum_{s=2}^{4} \operatorname{tr}(D_s^F)$, and the corresponding risk row is $\sum_{s=2}^{4} \overline{c}_s$. Adding ρ gives 65 no-risk or 85 with-risk parameters; both use $17(1 + 5^2) = 442$ moments. The VIF measures collinearity between the lagged-quantity response column and the remaining recovery design.

Table A.17 shows that the maintained quantity interpretation sharply selects no partial adjustment. Both the no-risk and with-risk estimates choose $\widehat{\rho} = 0$, and ρ is at the lower boundary in all 200 bootstrap draws. This is not a rank failure: the with-risk Jacobian has rank 85/85, and the lagged-quantity VIF is only 1.65. The first-week own sensitivity remains 5.049, with a 95% interval of [4.488,5.605]. Controlling for risk raises it by 0.011, with an interval of [0.006,0.015], and does so in every bootstrap draw. Most importantly, the weeks-two-through-four systematic sensitivity remains 8.167, with an interval of [4.592,10.727], and is positive in every draw.

The predetermined-state controls deliver the same result. The estimate of ρ remains zero

in every draw, first-week own sensitivity is 5.102, and the continuation trace is 6.899 with an interval of [4.147,9.575]. The cumulative-inventory mapping instead selects the imposed upper bound 0.95 in every draw. This is the accounting force described after (A.11), not evidence of slow preference adjustment. Even under this mapping, the continuation trace remains positive and precisely estimated.

	Week 1	Week 2	Week 3	Week 4
Panel A: Maintained level-state specification				
Identity component $a_{D,s}$	5.054	0.000	0.000	0.000
	(0.283)	(0.000)	(0.000)	(0.000)
Avg. stock own sensitivity $\overline{D}_{s,ii}$	5.04917	0.00059	0.00041	0.00066
	(0.28221)	(0.00015)	(0.00014)	(0.00019)
Systematic sensitivity $\mathrm{tr}(D_s^F)$	0.930	2.890	1.993	3.284
	(0.528)	(0.741)	(0.672)	(0.908)
Avg. own-variance loading $\overline{c}_s$	0.972	0.074	0.113	0.000
	(0.316)	(0.099)	(0.240)	(0.039)
Panel B: Four predetermined OFI and return lags				
Identity component $a_{D,s}$	5.107	0.000	0.000	0.000
	(0.307)	(0.000)	(0.000)	(0.000)
Avg. stock own sensitivity $\overline{D}_{s,ii}$	5.10189	0.00053	0.00038	0.00049
	(0.30699)	(0.00014)	(0.00014)	(0.00017)
Systematic sensitivity $\mathrm{tr}(D_s^F)$	1.089	2.579	1.870	2.450
	(0.567)	(0.681)	(0.678)	(0.847)
Avg. own-variance loading $\overline{c}_s$	0.904	0.114	0.039	0.000
	(0.283)	(0.083)	(0.261)	(0.046)

Table A.18. Weekly demand kernels with lagged quantity
This table reports the with-risk weekly estimates when ρ is estimated jointly with the four demand kernels. Panel A uses the maintained level-state specification. Panel B adds four predetermined OFI lags and four predetermined excess-return lags to the stock-level OFI and return LPs. Each second-stage estimate has 85 parameters—the preferred 84 demand parameters plus ρ—and 442 moments. Bootstrap standard errors from 200 twelve-week block-bootstrap draws are in parentheses. The PCA basis is held fixed across draws.

Table A.18 shows where the robustness result comes from. As in the main specification, the identity component is concentrated entirely in week one. The systematic diagonal terms, however, remain economically large in weeks two, three, and four under both first-stage designs. Controlling

for pre-shock history therefore changes the allocation across factor directions somewhat, but it does not turn the representative investor into a one-week investor. The first-week own-variance loading is positive and precisely separated from zero. Later risk loadings remain less precisely allocated, consistent with the main results.

Fixed ρ	RMSE	First-week own	Weeks 2–4 systematic	Weeks 2–4 risk
0.00	0.01321	5.049	8.167	0.188
0.25	0.02344	4.589	11.060	0.687
0.50	0.03468	4.129	14.527	1.213
0.75	0.04620	3.669	18.660	1.737
0.90	0.05318	3.393	21.166	2.052

Table A.19. Fixed-ρ profile
This table fixes ρ at the indicated value and re-estimates all 84 expected-return and risk coefficients in the maintained level-state specification. "First-week own" is the average stock own sensitivity, "Weeks 2–4 systematic" is $\sum_{s=2}^{4} \mathrm{tr}(D_s^F)$, and "Weeks 2–4 risk" is $\sum_{s=2}^{4} \overline{c}_s$. RMSE is computed from the full set of $17(1+5^2) = 442$ stacked trace and factor-block recovery moments through $h = 16$.

The fixed-ρ profile in Table A.19 makes the boundary estimate transparent. RMSE rises monotonically from 0.01321 at $\rho = 0$ to 0.02344 at $\rho = 0.25$ and 0.05318 at $\rho = 0.90$. Positive inertia does not eliminate continuation demand; it requires even larger systematic continuation coefficients. What it does undermine is the independently disciplined first-week own-elasticity anchor, which falls from 5.049 at $\rho = 0$ to 4.589 at $\rho = 0.25$ and 3.393 at $\rho = 0.90$. The data therefore distinguish lagged quantity from forward-looking systematic demand rather than trading one mechanically for the other.

L Stacked Recovery System

For a given horizon h, equation (A.4) is linear in the unknown coefficient matrices. Applying $\operatorname{vec}(AB) = (B^\top \otimes I)\operatorname{vec}(A)$ gives

$$\begin{aligned}\operatorname{vec}(\Psi_h^z) = ((\Psi_h^p)^\top \otimes I)\operatorname{vec}(A_0) - \sum_{s=1}^{S}((\Psi_{h+s}^p)^\top \otimes I)\operatorname{vec}(A_s) \\ - \sum_{s=1}^{S}\left((\Phi_{s,h}^d)^\top \otimes I\right)\operatorname{vec}(B_s) + \sum_{s=1}^{S}\left((\Omega_{s,h})^\top \otimes I\right)\operatorname{vec}(C_s).\end{aligned} \tag{A.12}$$

Let $\theta \in \Theta \subset \mathbb{R}^{K_\theta}$ denote the primitive parameter vector. We write $A_s(\theta)$ for example to denote A_s as a function of the parameters. The vector $g(\theta)$ maps the parameters to the signed vectorized coefficients as they appear in the recovery equation. If all matrices are unrestricted, we can parameterize these signed coefficients directly so that $g(\theta) = \theta$. Define

$$g_A(\theta) \equiv \left(\operatorname{vec}(A_0(\theta))^\top, -\operatorname{vec}(A_1(\theta))^\top, \ldots, -\operatorname{vec}(A_S(\theta))^\top\right)^\top.$$

$$g_B(\theta) \equiv \left(-\operatorname{vec}(B_1(\theta))^\top, \ldots, -\operatorname{vec}(B_S(\theta))^\top\right)^\top.$$

$$g_C(\theta) \equiv \left(\operatorname{vec}(C_1(\theta))^\top, \ldots, \operatorname{vec}(C_S(\theta))^\top\right)^\top, \qquad g(\theta) \equiv \left(g_A(\theta)^\top, g_B(\theta)^\top, g_C(\theta)^\top\right)^\top.$$

Also define

$$Y_H(\boldsymbol{\Psi}) \equiv \left(\operatorname{vec}(\Psi_0^z)^\top, \ldots, \operatorname{vec}(\Psi_H^z)^\top\right)^\top, \qquad X_H(\boldsymbol{\Psi}) \equiv \begin{bmatrix}\mathcal{X}_0 \\ \mathcal{X}_1 \\ \vdots \\ \mathcal{X}_H\end{bmatrix},$$

where the horizon-h block is

$$\mathcal{X}_h \equiv \begin{bmatrix}\mathcal{X}_h^A & \mathcal{X}_h^B & \mathcal{X}_h^C\end{bmatrix},$$

with

$$\mathcal{X}_h^A \equiv \left[(\Psi_h^p)^\top \otimes I,\ (\Psi_{h+1}^p)^\top \otimes I,\ \ldots,\ (\Psi_{h+S}^p)^\top \otimes I\right],$$

$$\mathcal{X}_h^B \equiv \left[(\Phi_{1,h}^d)^\top \otimes I,\ \ldots,\ (\Phi_{S,h}^d)^\top \otimes I\right], \qquad \mathcal{X}_h^C \equiv \left[(\Omega_{1,h})^\top \otimes I,\ \ldots,\ (\Omega_{S,h})^\top \otimes I\right].$$

Stacking (A.12) across horizons $h = 0, \ldots, H$ yields equation (22) in the main text.

The unrestricted benchmark is useful for counting. Each horizon contributes one $N \times N$ restriction, so stacking $h = 0, \ldots, H$ gives $(H+1)N^2$ scalar equations. If every coefficient entry is free, the unknown blocks contribute

$$A_0 : \ N^2, \qquad \{A_s\}_{s=1}^{S} : \ SN^2, \qquad \{B_s\}_{s=1}^{S} : \ SN^2, \qquad \{C_s\}_{s=1}^{S} : \ SNm,$$

where $C_s \in \mathbb{R}^{N \times m}$ and $m = N(N+1)/2$. The unrestricted system therefore contains

$$N^2(1+2S) + SNm$$

scalar unknowns, so a necessary condition for identification is

$$(H+1)N^2 \geq N^2(1+2S) + SNm,$$

or equivalently

$$H + 1 \geq 1 + 2S + \frac{Sm}{N} = 1 + 2S + \frac{S(N+1)}{2}.$$

This grows quickly with both horizon length and cross-sectional dimension, which is why the unrestricted system is mainly a benchmark.

Two special cases help calibrate the counting. If $B_s = C_s = 0$ and $A_s = 0$ for all $s \geq 1$, then only A_0 is unknown, giving N^2 parameters and requiring only $H + 1 \geq 1$. If $B_s = C_s = 0$ but $(A_1, \ldots, A_S)$ are retained, then the unknowns are $(A_0, \ldots, A_S)$ only, for a total of $(S+1)N^2$ parameters, and the necessary condition becomes $H + 1 \geq S + 1$.

Restrictions on the risk block can also lower the dimension substantially. As discussed in the text, if investors care only about each asset's own conditional variance and its conditional covariance with F common factors, then each risk matrix is $N \times (1+F)$ rather than $N \times m$. The unrestricted count then becomes

$$N^2(1+2S) + SN(1+F),$$

which is much smaller when $F \ll N$.

M A Sequence-Space Representation of the Portfolio Problem

This section derives the sequence-space representation of a general portfolio-choice problem. The environment is analogous to the one in Section 2, but here we start from the investor's optimization problem and show how its local solution can be written as the demand system used in the paper. The key step is to replace the history of shocks by the sequence of conditional forecasts of prices, dividends, and investor-specific disturbances.

Assets. Investors have access to N risky assets with price vector $P_t \in \mathbb{R}^N$, and a riskless asset with return $R_f \in \mathbb{R}$. For simplicity, we focus on the case where the riskless rate is constant. The risky assets pay out dividends given by $\mathcal{D}_t \in \mathbb{R}^N$. We assume that dividends are strictly positive at all times, so we can work with log-dividends, $d_t \equiv \log \mathcal{D}_t$, where the logarithm is applied elementwise. Similarly, we denote the vector of log prices as $p_t \equiv \log P_t$.

The economy is exposed to a vector of exogenous shocks $\nu_t \in \mathbb{R}^K$. The shocks are iid and normally distributed, i.e., $\nu_t \sim \mathcal{N}(\mathbf{0}_K, I_K)$. Prices and dividends are a function of the history of shocks: $p_t = p(\nu^t, t)$ and $d_t = d(\nu^t, t)$. We focus on a linear equilibrium where prices and dividends follow stationary processes given by

$$p_t = \overline{p} + \sum_{k=0}^{\infty} \Psi_{p,k} \nu_{t-k}, \qquad d_t = \overline{d} + \sum_{k=0}^{\infty} \Psi_{d,k} \nu_{t-k}. \tag{A.13}$$

The matrices $\Psi_{p,k}$ and $\Psi_{d,k}$ fully characterize the equilibrium price and dividend processes. In particular, they determine how a shock to the economy at date t affects prices and dividends at all future dates.

Frictions, preferences, and non-portfolio income. The elastic investor chooses consumption C_t, holdings of risky assets Q_t, and bond holdings B_t to maximize their expected utility. To capture the effect of financial frictions, we assume that portfolio holdings must satisfy the following constraints:

$$F(C_t, Q_t, B_t; \zeta_t) \geq 0, \tag{A.14}$$

where $F(\cdot) \in \mathbb{R}^M$ is a function that captures the effect of financial frictions. We assume that the function $F(\cdot)$ includes limits on portfolio positions to avoid Ponzi schemes, so the portfolio

problem is well-defined. The exogenous vector $\zeta_t = \zeta(\nu^t, t)$ collects the disturbances to the investor's problem, capturing, for instance, shifts in liquidity constraints or margin requirements.

We allow investors to receive utility from portfolio holdings, as in the literature on convenience yields or taste-based investment motives, like ESG. Formally, we allow the investor's utility to depend on portfolio holdings, $u(C_t, Q_t, B_t; \zeta_t)$.[16] The vector ζ_t here includes potential shocks to the taste for different assets, as well as the effect of discount rate shocks, as in valuation risk models (e.g., Albuquerque, Eichenbaum, Luo, and Rebelo 2016).

Finally, we assume that the investor receives non-portfolio income, $Y_t = Y(\zeta_t)$, capturing labor income or income from private businesses. Hence, the exogenous vector ζ_t collects the disturbances to financial constraints, preferences, and non-portfolio income.

We assume that ζ_t follows a stationary process given by

$$\zeta_t = \overline{\zeta} + \sum_{k=0}^{\infty} \Psi_{\zeta,k} \nu_{t-k}. \tag{A.15}$$

Investor's problem. Let $C = \{C_t\}_{t=0}^{\infty}$, $Q = \{Q_t\}_{t=0}^{\infty}$, and $B = \{B_t\}_{t=0}^{\infty}$ be the sequences of random variables representing consumption, portfolio holdings, and bond holdings. The investor chooses adapted sequences to solve

$$\max_{C,Q,B} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u(C_t, Q_t, B_t; \zeta_t) \right], \tag{A.16}$$

subject to

$$C_t + P_t^\top Q_t + R_f^{-1} B_t = (P_t + \mathcal{D}_t)^\top Q_{t-1} + B_{t-1} + Y_t,$$

the constraints (A.14), given the initial holdings (Q_{-1}, B_{-1}), the processes for prices and dividends defined in (A.13), and the process for the disturbances ζ_t given in (A.15).

Problem (A.16) corresponds to the sequence-space representation of the investor's problem. The investor takes as given the processes for prices, dividends, and disturbances to the investor's problem, and chooses the sequences C, Q, B that maximize their expected utility. This is in contrast to the recursive representation of the investor's problem, where prices and dividends are functions of an aggregate state variable, and the investor solves for a mapping from individual and aggregate state variables to portfolio holdings and consumption.

[16] For simplicity, we assume preferences are time-separable, but it is straightforward to generalize to recursive preferences.

From history of shocks to conditional forecasts. The solution to the sequence problem can be written as a function of the information available at date t, equivalently as a function of the history of shocks ν^t together with the inherited portfolio positions. We now show that, under a finite-order linear representation, the same information can be encoded by conditional forecasts.

Stack prices, dividends, and disturbances into a single centered vector

$$x_t = [(p_t - \overline{p})^\top, (d_t - \overline{d})^\top, (\zeta_t - \overline{\zeta})^\top]^\top.$$

It is convenient to assume that x_t follows an $MA(S)$ process:

$$x_t = \sum_{k=0}^{S} \Psi_{x,k} \nu_{t-k}.$$

The conditional expectation of future values of x_t is given by

$$\mathbb{E}_t[x_{t+s}] = \sum_{k=s}^{S} \Psi_{x,k} \nu_{t+s-k},$$

for $0 \le s \le S$, and $\mathbb{E}_t[x_{t+s}] = \mathbf{0}$ for $s > S$. Thus, in levels, forecasts beyond horizon S are equal to their unconditional means.

We can then write:

$$\mathbb{E}_t \underbrace{\begin{bmatrix} x_t \\ x_{t+1} \\ \vdots \\ x_{t+S-1} \\ x_{t+S} \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} \Psi_{x,0} & \Psi_{x,1} & \cdots & \Psi_{x,S-1} & \Psi_{x,S} \\ \Psi_{x,1} & \Psi_{x,2} & \cdots & \Psi_{x,S} & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \Psi_{x,S-1} & \Psi_{x,S} & \cdots & \mathbf{0} & \mathbf{0} \\ \Psi_{x,S} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{A_\nu} \underbrace{\begin{bmatrix} \nu_t \\ \nu_{t-1} \\ \vdots \\ \nu_{t-(S-1)} \\ \nu_{t-S} \end{bmatrix}}_{V_t}.$$

Assume that A_ν has full column rank. Then $A_\nu^\top A_\nu$ is nonsingular, so one possible left inverse is $L_\nu = (A_\nu^\top A_\nu)^{-1} A_\nu^\top$ and $V_t = L_\nu X_t$. The full-column-rank condition is the only requirement for the forecast vector to span the relevant shock history. Under this condition, $\nu_t, \nu_{t-1}, \ldots, \nu_{t-S}$ can be recovered from $x_t, \mathbb{E}_t[x_{t+1}], \ldots, \mathbb{E}_t[x_{t+S}]$. The vector X_t therefore summarizes all payoff-relevant

information in the finite-order linear economy. Equivalently, any object that is linear in the shock history can be written as a linear function of the forecast vector X_t.

The optimal policy can therefore be represented as

$$C_t = C(X_t, Q_{t-1}, B_{t-1}), \qquad Q_t = Q(X_t, Q_{t-1}, B_{t-1}), \qquad B_t = B(X_t, Q_{t-1}, B_{t-1}),$$

where the inherited positions are included because they determine cash on hand and the current budget set. This is the sequence-space counterpart of a recursive policy rule: the arguments are forecast paths rather than a specified Markov state.

The log-linearized demand system. We now derive the log-linearized demand system. Linearizing the portfolio rule around the stationary equilibrium gives

$$q_t = \sum_{s=0}^{S} G_s^p \mathbb{E}_t[p_{t+s}] + \sum_{s=0}^{S} G_s^d \mathbb{E}_t[d_{t+s}] + \sum_{s=0}^{S} G_s^\zeta \mathbb{E}_t[\zeta_{t+s}] + G^q q_{t-1} + G^b b_{t-1} + g_0,$$

where the coefficient matrices are derivatives of the optimal policy with respect to the forecast-path arguments and inherited positions. Constants and means are absorbed into g_0.

Collect the non-price terms that are not explicit arguments of the paper's demand system into a latent demand shifter,

$$\xi_t \equiv g_0 + G_0^d d_t + \sum_{s=0}^{S} G_s^\zeta \mathbb{E}_t[\zeta_{t+s}] + G^q q_{t-1} + G^b b_{t-1}.$$

Defining $A_0 \equiv -G_0^p$, $A_s \equiv G_s^p$ for $s \geq 1$, and $B_s \equiv G_s^d$ for $s \geq 1$, we obtain

$$q_t = -A_0 p_t + \sum_{s=1}^{S} A_s \mathbb{E}_t[p_{t+s}] + \sum_{s=1}^{S} B_s \mathbb{E}_t[d_{t+s}] + \xi_t,$$

which is the homoskedastic version of (1). The latent demand shifter ξ_t collects preference, constraint, and non-portfolio-income disturbances, current dividend cash-flow effects, and the dependence on inherited portfolio positions. In equilibrium, absent shocks to bond demand from inelastic investors, bond holdings are constant at their steady state, while lagged risky holdings are predetermined functions of past exogenous demand shocks.

N Nesting Different Demand Specifications

This appendix derives how our demand system (1) nests a wide range of models considered in the literature. We consider the static Koijen and Yogo (2019) system, intertemporal hedging demand (Merton, 1973; Campbell and Viceira, 2002), models with belief distortions (Gabaix, 2019; Bordalo et al., 2018) and dispersed information (Grossman and Stiglitz, 1980), and preferred-habitat no-arbitrage specifications (Vayanos and Vila, 2021). Each subsection linearizes a canonical model and maps it into our notation, making explicit the implied restrictions on (A_0, A_s, B_s, C_s) and on the latent demand shifter ξ_t.

N.1 Static models

This is the $S = 0$ case of (1): demand depends only on current prices, so $A_s = B_s = C_s = 0$ for all $s \geq 1$, as in (6).

Koijen and Yogo (2019). We consider a simplified version of their demand system that abstracts from observed characteristics other than the latent shifter $\Lambda_{j,t}$:

$$\frac{\omega_{j,t}}{\omega_{0,t}} = \exp\left(\beta_0 + \beta_1 p_{j,t}\right) \Lambda_{j,t} \equiv \delta_{j,t},$$

for assets $j = 1, \ldots, N$, where $\omega_{j,t}$ denotes the portfolio weight of asset j at time t, and $\Lambda_{j,t}$ is the latent demand shifter. Asset $j = 0$ is the outside asset, whose price we normalize to one, so $p_{0,t} = 0$; in our setting, we take it to be the risk-free asset with constant price.

The portfolio weight of asset j is given by the multinomial logit formula:

$$\omega_{j,t} = \frac{\delta_{j,t}}{1 + \sum_{i=1}^{N} \delta_{i,t}}.$$

Given the portfolio weight, we can solve for the quantity of asset j held $Q_{j,t}$ using the expression: $\omega_{j,t} \equiv \frac{P_{j,t} Q_{j,t}}{W_t (1 - C_t)}$, where $P_{j,t}$ denotes the price of asset j in levels, W_t denotes the wealth of the investor, and C_t denotes the consumption-wealth ratio. For simplicity, we assume the consumption-wealth ratio is constant in this static model.

We consider a linearization of the demand system around the approximation point $(\overline{P}, \overline{Q}, \overline{W}, \overline{C})$. We assume that the price and wealth are stationary, so that the approximation point is time-invariant.

It is straightforward to generalize the results to the case where prices and wealth are scaled by a stochastic trend, so normalized prices and wealth are stationary.

Linearizing the demand for asset $j = 0$ gives:

$$q_{0,t} = \overline{q}_0 - \sum_{j=1}^{N} \frac{\overline{\delta}_j}{1 + \sum_{i=1}^{N} \overline{\delta}_i} (\beta_1 p_{j,t} + \log \Lambda_{j,t}) + w_t - \overline{w},$$

where $\overline{q}_0$ collects the constant terms.

Abstracting from dividends, the wealth is given by $W_t = \sum_{j=0}^{N} P_{j,t} Q_{j,t-1}$. Linearizing this expression, we obtain:

$$w_t = \overline{w} + \sum_{j=0}^{N} \frac{\overline{P}_j \overline{Q}_j}{\overline{W}} \left[(p_{j,t} - \overline{p}_j) + (q_{j,t-1} - \overline{q}_j) \right]$$

where $\overline{W}$ is the approximation point for wealth. Note that the $j = 0$ term contributes only through $q_{0,t-1}$, since $p_{0,t} = 0$.

Combining the expressions above, we obtain:

$$q_{0,t} = \overline{q}_0 + \chi_0^\top p_t + \xi_{0,t},$$

where $p_t = (p_{1,t}, \ldots, p_{N,t})^\top$, and the vector χ_0 and shifter $\xi_{0,t}$ are given element by element by

$$(\chi_0)_j = \frac{\overline{P}_j \overline{Q}_j}{\overline{W}} - \frac{\overline{\delta}_j \beta_1}{1 + \sum_{i=1}^{N} \overline{\delta}_i}, \qquad j = 1, \ldots, N,$$

$$\xi_{0,t} = \sum_{j=0}^{N} \frac{\overline{P}_j \overline{Q}_j}{\overline{W}} q_{j,t-1} - \sum_{j=1}^{N} \frac{\overline{\delta}_j}{1 + \sum_{i=1}^{N} \overline{\delta}_i} \log \Lambda_{j,t}.$$

where the remaining constant terms are absorbed into $\overline{q}_0$. Each element of χ_0 combines a wealth effect, as a higher price $p_{j,t}$ raises wealth and hence the demand for the outside asset, and a substitution effect, as a higher price changes the relative attractiveness of asset j.

Taking logs of the demand for asset $j \in \{1, \ldots, N\}$, and using $\omega_{j,t}/\omega_{0,t} = P_{j,t} Q_{j,t} / (P_{0,t} Q_{0,t})$

with $p_{0,t} = 0$, gives the exact relation:

$$q_{j,t} = q_{0,t} + \beta_0 + (\beta_1 - 1)p_{j,t} + \log \Lambda_{j,t}.$$

Combining the expressions above, we obtain:

$$q_t = -A_0 p_t + \xi_t, \tag{A.17}$$

where

$$A_0 = (1 - \beta_1)I_N - \mathbf{1}_N \times \chi_0^\top, \qquad \xi_{j,t} = \overline{q}_0 + \beta_0 + \xi_{0,t} + \log \Lambda_{j,t}.$$

Differentiating (A.17) gives the own- and cross-price elasticities

$$\frac{\partial q_{i,t}}{\partial p_{j,t}} = -(1 - \beta_1)\mathbf{1}_{\{i=j\}} - \chi_{0,j}.$$

Thus the cross-price elasticity of asset i with respect to $p_{j,t}$ is $-\chi_{0,j}$ for $i \neq j$, while the own-price elasticity is $-(1 - \beta_1) - \chi_{0,i}$. The matrix A_0 therefore encodes the contemporaneous own- and cross-price responses, while ξ_t collects the latent demand shifters $\log \Lambda_{j,t}$, lagged holdings, and constants.

N.2 Intertemporal hedging demands

We next nest intertemporal hedging demand from the Campbell and Viceira (2002) log-linearization of Merton (1973). In the homoskedastic case, the conditional covariance matrix Σ is constant, so demand responds to the path of expected returns while higher moments are fixed; in (1), this corresponds to $C_s = 0$ for all s (or, equivalently, to risk terms that enter only through constants absorbed into ξ_t). The coefficients $(A_s, B_s)_{s=1}^S$ are unrestricted beyond what is implied by the return-based representation below.

We focus on the case in which expected returns are affine in a K-dimensional state X_t,

$$\mathbb{E}_t[r_{t+1}] = \mu_r + \Psi X_t,$$

where μ_r is the unconditional mean of returns, Ψ is an $N \times K$ matrix, and the state follows the

VAR(1) process

$$X_t = \Phi_1 X_{t-1} + \epsilon_t, \tag{A.18}$$

where ϵ_t is a K-dimensional vector of white-noise shocks.

Campbell and Viceira (2002) linearize the portfolio share equation as

$$\alpha_t = \underbrace{\frac{1}{\gamma}\Sigma^{-1}\left(\mathbb{E}_t[r_{t+1}] + \frac{\sigma^2}{2}\right)}_{\text{myopic demand}} + \underbrace{A_{\alpha,0} + A_{\alpha,X}X_t}_{\text{hedging demand}},$$

where σ^2 denotes the vector of diagonal elements of Σ, so the Jensen correction is part of the myopic demand and hedging demand vanishes in the log-utility case.

Portfolio shares are therefore affine in X_t. Equivalently, they can be written as a function of the sequence of expected returns. Since

$$\mathbb{E}_t[r_{t+s}] = \mu_r + \Psi\Phi_1^{s-1}X_t,$$

we stack expected returns over horizons $1, \ldots, S$:

$$\mathcal{R}_t \equiv \begin{bmatrix} \mathbb{E}_t[r_{t+1}] \\ \vdots \\ \mathbb{E}_t[r_{t+S}] \end{bmatrix} = \mu_{\mathcal{R}} + \widetilde{\Psi}X_t,$$

where $\mu_{\mathcal{R}} = \mathbf{1}_S \otimes \mu_r$ and $\widetilde{\Psi} = [\Psi^\top, (\Psi\Phi_1)^\top, \ldots, (\Psi\Phi_1^{S-1})^\top]^\top$ is an $NS \times K$ matrix. If S is large enough that $\operatorname{rank}(\widetilde{\Psi}) = K$, the state can be recovered from the expected-return sequence,

$$X_t = (\widetilde{\Psi}^\top\widetilde{\Psi})^{-1}\widetilde{\Psi}^\top(\mathcal{R}_t - \mu_{\mathcal{R}}).$$

If expected returns alone do not span the state—for example, because volatility also moves with X_t—the joint path of expected returns and risk would be required instead.

Substituting into the portfolio-share equation and defining

$$\tilde{A}_{\alpha,0} \equiv A_{\alpha,0} - A_{\alpha,X}(\widetilde{\Psi}^\top\widetilde{\Psi})^{-1}\widetilde{\Psi}^\top\mu_{\mathcal{R}},$$

we obtain

$$\alpha_t = \tilde{A}_{\alpha,0} + \frac{1}{\gamma}\Sigma^{-1}\left(\mathbb{E}_t[r_{t+1}] + \frac{\sigma^2}{2}\right) + A_{\alpha,X}(\widetilde{\Psi}^\top\widetilde{\Psi})^{-1}\widetilde{\Psi}^\top\mathcal{R}_t.$$

Defining the consecutive $N \times N$ blocks of $A_{\alpha,X}(\widetilde{\Psi}^\top\widetilde{\Psi})^{-1}\widetilde{\Psi}^\top$ as $\{\tilde{C}_s\}_{s=1}^S$, we obtain

$$\alpha_t = \tilde{A}_{\alpha,0} + \frac{1}{\gamma}\Sigma^{-1}\left(\mathbb{E}_t[r_{t+1}] + \frac{\sigma^2}{2}\right) + \sum_{s=1}^{S}\tilde{C}_s\mathbb{E}_t[r_{t+s}].$$

The same linearization also implies $C_t = A_{C,0} + A_{C,X}X_t$, and hence that C_t can be expressed as a function of $\mathcal{R}_t$. Using $Q_{j,t} = \alpha_{j,t}(1 - C_t)W_t/P_{j,t}$ and linearizing around the steady state as in the static subsection above gives

$$q_t = -A_0^w p_t + \sum_{s=1}^{S} D_s\mathbb{E}_t[r_{t+s}] + \xi_t,$$

where $A_0^w = I_N - \mathbf{1}_N\overline{s}^\top$, with $\overline{s} = (\overline{P}_1\overline{Q}_1/\overline{W}, \ldots, \overline{P}_N\overline{Q}_N/\overline{W})^\top$ the vector of steady-state wealth shares, and where D_s combine the contributions of α_t and C_t. The shifter ξ_t collects lagged holdings and constants from the wealth expansion; since the risk-free rate is constant, expected returns and expected excess returns differ only by constants absorbed into ξ_t. This is the multivariate generalization of the return-based demand system (3), augmented by the contemporaneous price term $-A_0^w p_t$ through the wealth channel.

Under log utility, hedging demand vanishes and the consumption-wealth ratio is constant, so $D_s = 0$ for all $s > 1$ and demand depends only on the current expected return. With Epstein–Zin preferences, intertemporal links can remain even when the EIS is one (through hedging) or when risk aversion is one (through the consumption-wealth ratio). The sequence-space representation therefore nests Merton hedging while allowing the econometrician to work directly with the path $\{\mathbb{E}_t[r_{t+s}]\}_{s=1}^S$ without observing X_t.

N.3 Behavioral models and informational frictions

The preceding subsections nest canonical demand systems under full-information rational expectations. However, a large literature studies models with belief distortions, where expectations

deviate from the rational benchmark. Our demand system can accommodate a range of behavioral models, as well as rational models in which investors face informational frictions.

Starting from (1), suppose the investor's demand depends on subjective rather than objective expected paths. To capture belief distortions and informational frictions jointly, we write demand in terms of the *subjective* or perceived paths of prices, dividends, and risk:

$$q_t = -A_0\, p_t + \sum_{s=1}^{S} \hat{A}_s\, \hat{\mu}^p_{t,t+s} + \sum_{s=1}^{S} \hat{B}_s\, \hat{\mu}^d_{t,t+s} - \sum_{s=1}^{S} \hat{C}_s\, \hat{v}^r_{t,t+s}, \tag{A.19}$$

where $\hat{\mu}^p_{t,t+s}$ and $\hat{\mu}^d_{t,t+s}$ are the investor's perceived paths of prices and dividends at horizon s, $\hat{v}^r_{t,t+s}$ is the perceived conditional return risk at horizon s, and $\hat{A}_s$,$\hat{B}_s \in \mathbb{R}^{N\times N}$ and $\hat{C}_s \in \mathbb{R}^{N\times m}$ are the subjective demand coefficients.

We assume that the subjective values are related to the objective ones through the following relationship:

$$\hat{\mu}^p_{t,t+s} = M^p_s\, \mu^p_{t,t+s} + \xi^p_{t,t+s}, \qquad \hat{\mu}^d_{t,t+s} = M^d_s\, \mu^d_{t,t+s} + \xi^d_{t,t+s}, \qquad \hat{v}^r_{t,t+s} = M^r_s\, v^r_{t,t+s} + \xi^r_{t,t+s}, \tag{A.20}$$

where M^p_s,$M^d_s \in \mathbb{R}^{N\times N}$, $M^r_s \in \mathbb{R}^{m\times m}$, and $\xi^p_{t,t+s}$,$\xi^d_{t,t+s} \in \mathbb{R}^N$ and $\xi^r_{t,t+s} \in \mathbb{R}^m$ are belief shocks, dated t and indexed by the horizon $t+s$ they refer to. Other latent demand shifters can be subsumed in these shocks or appear after substituting (A.20) into (A.19).

We show next that a range of belief distortions and informational frictions proposed in the literature fit this general framework.

Behavioral inattention. In the behavioral inattention model of Gabaix (2019) and Gabaix (2023), investors solve a dynamic problem with a perceived law of motion for the state variables that differs from the true one. Gabaix (2023) represents this distortion in state space: if the true law of motion is (A.18), the investor perceives

$$X_t = M_X \Phi_1 X_{t-1} + M_\epsilon \epsilon_t,$$

where M_X is a $K \times K$ matrix and M_ϵ is a $K \times K$ diagonal matrix of attention coefficients. In the baseline Gabaix representation, inattention is deterministic and operates through (M_X,M_ϵ); we allow additional idiosyncratic belief shocks $\xi^{p,d,r}_{t,t+s}$ in (A.20) on top of this attenuation.

Under the hedging setup above, $\mu^p_{t,t+s}$, $\mu^d_{t,t+s}$, and $v^r_{t,t+s}$ are affine in $\Phi_1^{s-1} X_t$, so attenuating the

state transition and shock loading maps into horizon-dependent matrices M_s^p, M_s^d, M_s^r in (A.20). Alternatively, one can view (A.20) as the direct belief distortions for an investor who solves a sequential problem. Current prices are observed, so A_0 is unchanged.

Substituting (A.20) into (A.19) yields the objective demand system

$$q_t = -A_0\, p_t + \sum_{s=1}^{S} A_s\, \mu_{t,t+s}^p + \sum_{s=1}^{S} B_s\, \mu_{t,t+s}^d - \sum_{s=1}^{S} C_s\, v_{t,t+s}^r + \xi_t,$$

with composite coefficients

$$A_s = \hat{A}_s M_s^p, \qquad B_s = \hat{B}_s M_s^d, \qquad C_s = \hat{C}_s M_s^r,$$

and shifter

$$\xi_t = \sum_{s=1}^{S} \left(\hat{A}_s \xi_{t,t+s}^p + \hat{B}_s \xi_{t,t+s}^d - \hat{C}_s \xi_{t,t+s}^r \right). \tag{A.21}$$

Two implications follow. First, only the composite coefficients (A_s, B_s, C_s) are identified from data on objective expectations: a behavioral investor who misperceives future paths is observationally equivalent to a rational investor with different intertemporal elasticities to prices, dividends, and risk. For counterfactuals that move entire expected paths of prices, dividends, and risk, these composites are the relevant objects, so the recovery analysis applies unchanged. Second, when belief shocks $\xi_{t,t+s}^{p,d,r}$ are present, (A.21) provides a micro-foundation for the latent demand shifter ξ_t.

Diagnostic expectations. Diagnostic expectations (Bordalo et al., 2018) provide a second example in which behavioral beliefs can be mapped into our sequential demand representation. The key difference relative to behavioral inattention is that diagnostic beliefs depend not only on the current forecast, but also on a reference forecast formed in the past. To keep the equilibrium finite dimensional, we follow the "shadow rational-expectations" implementation in Bianchi, Ilut, and Saijo (2024): investors form diagnostic forecasts by distorting forecasts from a shadow rational-expectations economy rather than recursively distorting the forecasts generated by the diagnostic economy itself.

In the shadow economy, the primitive state follows (A.18). Under shadow rational expectations,

any demand-relevant object is affine in the state. For concreteness, consider expected returns,

$$\mathbb{E}_t^{RE}[r_{t+s}] = \mu_r + \Psi\Phi_1^{s-1}X_t.$$

To keep the notation transparent, consider the case in which the reference forecast is formed one period ago. Diagnostic expectations distort the shadow forecast relative to this recent reference forecast:

$$\widehat{\mathbb{E}}_t[r_{t+s}] = \mathbb{E}_t^{RE}[r_{t+s}] + \theta\left(\mathbb{E}_t^{RE}[r_{t+s}] - \mathbb{E}_{t-1}^{RE}[r_{t+s}]\right), \tag{A.22}$$

where $\theta \geq 0$ measures the strength of diagnostic overreaction; when $\theta = 0$, beliefs coincide with rational expectations. Since the anchor is the shadow rational-expectations law of motion, the lagged reference forecast is

$$\mathbb{E}_{t-1}^{RE}[r_{t+s}] = \mu_r + \Psi\Phi_1^{s}X_{t-1}.$$

Substituting into (A.22) gives

$$\widehat{\mathbb{E}}_t[r_{t+s}] = \mu_r + (1+\theta)\Psi\Phi_1^{s-1}X_t - \theta\Psi\Phi_1^{s}X_{t-1}. \tag{A.23}$$

Thus diagnostic beliefs expand the investor's relevant state from X_t to

$$\mathcal{S}_t^D \equiv \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix}.$$

The same logic applies to prices, dividends, and risk. If the shadow rational-expectations forecast of any demand-relevant object y_{t+s} is

$$\mathbb{E}_t^{RE}[y_{t+s}] = \mu_y + \Psi_y\Phi_1^{s-1}X_t,$$

then the diagnostic forecast is

$$\widehat{\mathbb{E}}_t[y_{t+s}] = \mu_y + (1+\theta)\Psi_y\Phi_1^{s-1}X_t - \theta\Psi_y\Phi_1^{s}X_{t-1}, \tag{A.24}$$

which is affine in $\mathcal{S}_t^D = (X_t^\top, X_{t-1}^\top)^\top$. The objective equilibrium expectations derived below are affine in the same expanded state, with potentially different coefficients. Once the diagnostic

equilibrium is solved and $\mathcal{S}_t^D$ is spanned by the objective expected-return sequence, subjective and objective paths can therefore be mapped into each other as in (A.20).

Consider now the portfolio-share representation. After solving the diagnostic-expectations equilibrium, let $\mathbb{E}_t[r_{t+s}]$ denote the *objective* expected return in that equilibrium—the quantity that appears in market clearing and recovery, not the shadow-RE benchmark $\mathbb{E}_t^{RE}[r_{t+s}]$ nor the diagnostic belief $\widehat{\mathbb{E}}_t[r_{t+s}]$. Prices and these objective expected returns are affine in the expanded state:

$$\mathbb{E}_t[r_{t+s}] = \mu_s^D + \Psi_s^D \mathcal{S}_t^D, \qquad s = 1, \ldots, S.$$

The investor's portfolio share is also affine in the same state,

$$\alpha_t = A_{\alpha,0}^D + A_{\alpha,S}^D \mathcal{S}_t^D,$$

with the myopic mean–variance component absorbed into the intercept $A_{\alpha,0}^D$. Stack the objective expected returns in the diagnostic equilibrium,

$$\mathcal{R}_t^D \equiv \begin{bmatrix} \mathbb{E}_t[r_{t+1}] \\ \vdots \\ \mathbb{E}_t[r_{t+S}] \end{bmatrix} = \mu_{\mathcal{R}}^D + \widetilde{\Psi}_D \mathcal{S}_t^D, \qquad \widetilde{\Psi}_D \equiv \begin{bmatrix} \Psi_1^D \\ \vdots \\ \Psi_S^D \end{bmatrix}.$$

Stack diagnostic return beliefs the same way,

$$\widehat{\mathcal{R}}_t^D \equiv \begin{bmatrix} \widehat{\mathbb{E}}_t[r_{t+1}] \\ \vdots \\ \widehat{\mathbb{E}}_t[r_{t+S}] \end{bmatrix} = \widehat{\mu}_{\mathcal{R}}^D + \widehat{\Psi}_D \mathcal{S}_t^D,$$

where $\widehat{\Psi}_D$ collects the diagnostic coefficients from (A.23). If S is large enough that $\operatorname{rank}(\widetilde{\Psi}_D) = \dim(\mathcal{S}_t^D)$, then the expanded state can be recovered from the objective expected-return sequence:

$$\mathcal{S}_t^D = (\widetilde{\Psi}_D^\top \widetilde{\Psi}_D)^{-1} \widetilde{\Psi}_D^\top (\mathcal{R}_t^D - \mu_{\mathcal{R}}^D),$$

and diagnostic beliefs can be written as linear functions of objective equilibrium expectations:

$$\widehat{\mathcal{R}}_t^D = \widehat{\mu}_{\mathcal{R}}^D + \widehat{\Psi}_D (\widetilde{\Psi}_D^\top \widetilde{\Psi}_D)^{-1} \widetilde{\Psi}_D^\top (\mathcal{R}_t^D - \mu_{\mathcal{R}}^D). \tag{A.25}$$

The same logic applies to prices, dividends, and risk through (A.24). Horizon-by-horizon, (A.25) delivers the matrices (M_s^p, M_s^d, M_s^r) in (A.20); unlike behavioral inattention, they are equilibrium objects rather than direct attention parameters on the state law of motion, and they generally couple horizons. Substituting the recovered state into the portfolio-share equation and defining

$$\widetilde{A}^D_{\alpha,0} \equiv A^D_{\alpha,0} - A^D_{\alpha,S}(\widetilde{\Psi}_D^\top \widetilde{\Psi}_D)^{-1}\widetilde{\Psi}_D^\top \mu^D_{\mathcal{R}},$$

we obtain

$$\alpha_t = \widetilde{A}^D_{\alpha,0} + A^D_{\alpha,S}(\widetilde{\Psi}_D^\top \widetilde{\Psi}_D)^{-1}\widetilde{\Psi}_D^\top \mathcal{R}^D_t.$$

Defining the consecutive $N \times N$ blocks of

$$A^D_{\alpha,S}(\widetilde{\Psi}_D^\top \widetilde{\Psi}_D)^{-1}\widetilde{\Psi}_D^\top$$

as $\{\widetilde{D}^D_s\}_{s=1}^S$, we obtain

$$\alpha_t = \widetilde{A}^D_{\alpha,0} + \sum_{s=1}^{S} \widetilde{D}^D_s \mathbb{E}_t[r_{t+s}].$$

Therefore, diagnostic expectations nest our sequential demand system along the same return-spanning route as intertemporal hedging. The behavioral distortion changes the equilibrium mapping from states to prices and returns and expands the relevant state by introducing one lagged memory state, but once the diagnostic equilibrium is solved, portfolio demand can be written as a linear function of the sequence of objective expected future returns. The same spanning argument applies to prices, dividends, and risk through (A.24); here we display the return-based representation for brevity. Applying the same log-linearization from portfolio shares to quantities as in the hedging subsection gives

$$q_t = -A^w_0 p_t + \sum_{s=1}^{S} D^D_s \mathbb{E}_t[r_{t+s}] + \xi^D_t,$$

where A^w_0 is the wealth-channel price coefficient defined above, D^D_s are the log-linearized counterparts of $\widetilde{D}^D_s$ and incorporate both intertemporal hedging motives and the diagnostic belief distortion, and ξ^D_t collects the residual terms from the log-linearization of portfolio shares

into quantities—including dependence on lagged holdings—as well as any forecast wedge if the econometrician's information set is coarser than $\mathcal{S}_t^D$.

Noisy rational expectations. Noisy-information models provide another route through which our sequential demand system arises. Unlike the behavioral inattention and diagnostic-expectations examples above, investors remain fully rational conditional on their information set. The departure from full-information rational expectations comes from the fact that investors do not directly observe the state vector driving expected returns.

The true state follows (A.18), and investors observe the noisy signal

$$Y_t = HX_t + \eta_t,$$

where H maps the state into signals and η_t is a white-noise information shock, independent of ϵ_t.

In a general noisy rational-expectations equilibrium, prices themselves convey information (Grossman and Stiglitz, 1980). In a linear equilibrium, observing the price system is informationally equivalent to observing the endogenous signal

$$\tilde{X}_t = GX_t + u_t,$$

where u_t is the noise that prevents prices from being fully revealing—outside supply, for instance—and the matrix G is determined in equilibrium.

Under linear-Gaussian dynamics, investors form the filtered estimate

$$\hat{X}_t \equiv \mathbb{E}[X_t \mid Y^t, \tilde{X}^t],$$

which evolves recursively according to the Kalman filter,

$$\hat{X}_t = K_Y Y_t + K_{\tilde{X}} \tilde{X}_t + K_X \hat{X}_{t-1},$$

for matrices $(K_Y, K_{\tilde{X}}, K_X)$ given by the steady-state Kalman gain. Because the signal $\tilde{X}_t$ is endogenous, the gains are determined as a fixed point: beliefs shape demand, demand shapes the pricing rule, and the pricing rule shapes the informativeness of prices.

Rather than describing the equilibrium from the perspective of an outside observer who sees

the true state, we work directly with the investor-observable state

$$\mathcal{S}_t^{NR} \equiv \begin{bmatrix} \hat{X}_t \\ \tilde{X}_t \end{bmatrix},$$

which is a sufficient statistic for the investor's forecasts: $\hat{X}_t$ summarizes beliefs about fundamentals, while $\tilde{X}_t$ pins down the current price and the investor's estimate of the current noise.

The investor's conditional expectations of future returns are then affine in this state:

$$\widehat{\mathbb{E}}_t[r_{t+s}] \equiv \mathbb{E}[r_{t+s} \mid Y^t, \tilde{X}^t] = \mu_s^{NR} + \Psi_s^{NR}\mathcal{S}_t^{NR}, \qquad s = 1, \ldots, S.$$

Likewise, portfolio shares are affine in the same state,

$$\alpha_t = A_{\alpha,0}^{NR} + A_{\alpha,S}^{NR}\mathcal{S}_t^{NR},$$

with the myopic mean–variance component absorbed into $A_{\alpha,0}^{NR}$. Stacking the investor's expected returns gives

$$\mathcal{R}_t^{NR} = \mu_{\mathcal{R}}^{NR} + \widetilde{\Psi}_{NR}\mathcal{S}_t^{NR},$$

where

$$\widetilde{\Psi}_{NR} \equiv \begin{bmatrix} \Psi_1^{NR} \\ \vdots \\ \Psi_S^{NR} \end{bmatrix}.$$

If $\widetilde{\Psi}_{NR}$ has full column rank, the investor-observable state can be recovered from the investor's expected-return sequence,

$$\mathcal{S}_t^{NR} = (\widetilde{\Psi}_{NR}^{\top}\widetilde{\Psi}_{NR})^{-1}\widetilde{\Psi}_{NR}^{\top}(\mathcal{R}_t^{NR} - \mu_{\mathcal{R}}^{NR}),$$

and portfolio demand can be written as a linear function of $\widehat{\mathbb{E}}_t[r_{t+s}]$ using the same spanning argument as above.

It is natural to assume that the econometrician does not observe the full information set $\{Y^t, \tilde{X}^t\}$ used by investors. For instance, the econometrician may not observe all private signals, filtered beliefs, or informational variables that enter investors' forecasts. Let $\mathcal{I}_t^E \subset \{Y^t, \tilde{X}^t\}$ denote the

information set available to the econometrician and define

$$\mathbb{E}_t[r_{t+s}] \equiv \mathbb{E}[r_{t+s} \mid \mathcal{I}_t^E]$$

for the econometrician's forecast of future returns. Define the forecast wedge

$$\xi_{t,t+s}^{NR} \equiv \widehat{\mathbb{E}}_t[r_{t+s}] - \mathbb{E}[r_{t+s} \mid \mathcal{I}_t^E], \tag{A.26}$$

so that $\widehat{\mathbb{E}}_t[r_{t+s}] = \mathbb{E}_t[r_{t+s}] + \xi_{t,t+s}^{NR}$. By the law of iterated expectations—using that $\mathcal{I}_t^E$ is nested in the investor's information set—$\mathbb{E}[\xi_{t,t+s}^{NR} \mid \mathcal{I}_t^E] = 0$: the wedge is orthogonal to the econometrician's information set. This orthogonality delivers the exclusion restriction in Assumption 1, provided the identified shocks are measurable with respect to $\mathcal{I}_t^E$. When $\mathcal{I}_t^E = \{Y^t, \tilde{X}^t\}$, the wedge vanishes and the representation collapses to standard sequential demand with no latent shifter.

Relative to the econometrician's measured expectations, (A.20) therefore takes the special case $M_s^p = M_s^d = M_s^r = I$; the entire departure from measured expectations operates through the forecast wedges $\xi_{t,t+s}^{NR}$, not through attenuation of objective paths. Substituting (A.26) into the portfolio-share equation and defining

$$\widetilde{A}_{\alpha,0}^{NR} \equiv A_{\alpha,0}^{NR} - A_{\alpha,S}^{NR}(\widetilde{\Psi}_{NR}^\top \widetilde{\Psi}_{NR})^{-1}\widetilde{\Psi}_{NR}^\top \mu_{\mathcal{R}}^{NR},$$

we obtain

$$\alpha_t = \widetilde{A}_{\alpha,0}^{NR} + \sum_{s=1}^{S} \widetilde{D}_s^{NR}\, \mathbb{E}_t[r_{t+s}] + \xi_t^{NR}, \qquad \xi_t^{NR} \equiv \sum_{s=1}^{S} \widetilde{D}_s^{NR}\, \xi_{t,t+s}^{NR},$$

where the consecutive $N \times N$ blocks of $A_{\alpha,S}^{NR}(\widetilde{\Psi}_{NR}^\top \widetilde{\Psi}_{NR})^{-1}\widetilde{\Psi}_{NR}^\top$ are $\{\widetilde{D}_s^{NR}\}_{s=1}^S$. Thus ξ_t^{NR} is not a behavioral belief distortion: it is an omitted-information term arising because investors condition on a richer information set than the researcher. Applying the same log-linearization from portfolio shares to quantities as in the hedging and diagnostic subsections gives

$$q_t = -A_0^w p_t + \sum_{s=1}^{S} D_s^{NR}\mathbb{E}_t[r_{t+s}] + \xi_t^{NR},$$

where D_s^{NR} are the log-linearized counterparts of $\widetilde{D}_s^{NR}$.

This example illustrates an important source of latent demand. If the econometrician observed the same information set as investors, the latent demand component would disappear, because

all relevant information would already be incorporated into measured expected returns; it arises only because the econometrician observes a coarser information set than investors. From the investors' perspective, $\widehat{\mathbb{E}}_t[r_{t+s}]$ is the relevant expectation; from the econometrician's perspective, the component not spanned by $\mathbb{E}_t[r_{t+s}]$ appears as an unobserved demand shifter.

Together, these examples show that the sequential demand system (1) and the recovery exercise in Section 5 apply once expectations enter through their measured, objective counterparts. Behavioral inattention reshapes the identified intertemporal elasticities through the matrices M_s^p, M_s^d, M_s^r but leaves the recovery machinery unchanged. Diagnostic expectations change the equilibrium mapping from states to prices and returns while preserving the return-spanning representation. Noisy information leaves the structural coefficients on measured expectations unchanged but introduces ξ_t^{NR} whenever the econometrician conditions on a coarser information set than investors. In all three cases, Assumption 1 remains the relevant exclusion restriction whenever ξ_t is not directly shifted by the identifying outside-demand shock.

N.4 Preferred-habitat models

Finally, we nest preferred-habitat no-arbitrage models (Vayanos and Vila, 2021). In (1), this is the $S = 1$ mean–variance case with $C_s = 0$, $A_1 = \beta A_0$, and $A_s = B_s = 0$ for $s \geq 2$: a myopic arbitrageur responds only to next-period expected prices and dividends while second moments are constant. Habitat investors enter either through outside holdings or through an additive contribution to A_0; the latent shifter ξ_t collects linearization constants and does not represent habitat demand, which is absorbed into z_t or A_0^{hab}.

Vayanos and Vila (2021) develop the leading example for the term structure, where habitat investors prefer specific bond maturities and competitive arbitrageurs absorb residual demand; the same logic applies whenever a price-elastic mean–variance arbitrageur clears against an inelastic or asset-specific clientele. There are N assets, habitat demand that may be price-inelastic or price-elastic, and a representative arbitrageur who maximizes $\mathbb{E}_t[W_{t+1}] - \frac{\gamma}{2}\mathbb{V}\text{ar}_t(W_{t+1})$ with risk aversion $\gamma > 0$ and conditional covariance $\Sigma \succ 0$. The arbitrageur's first-order condition delivers (7)–(8), so A_0^{arb} is proportional to Σ^{-1} up to the diagonal scaling in (8).[17] When habitat demand is price-inelastic, it is subsumed into the outside-holdings vector Z_t defined in Section 2.1; when it is

[17] Vayanos and Vila (2021) work in continuous time with the analogous instantaneous mean–variance objective $\mathbb{E}_t[dW_t] - \frac{a}{2}\mathbb{V}\text{ar}_t(dW_t)$ in place of the one-period objective used here. Both formulations deliver the same mean–variance demand: arbitrageur holdings proportional to Σ^{-1} times the conditional risk premium.

price-elastic, the aggregate own-price coefficient decomposes as $A_0 = A_0^{\text{arb}} + A_0^{\text{hab}}$. Market clearing then reduces to (2).

Combined with the arbitrageur's pricing condition—expected returns move with $\gamma\Sigma$ times the residual portfolio the arbitrageur must hold—this delivers the familiar mapping from supply shocks in z_t to risk premia in the affine equilibrium of Vayanos and Vila (2021). No-arbitrage is therefore not in tension with a demand-system representation: the structural coefficients (A_0, A_1) encode the arbitrageur's effective price of risk, with γ and Σ determining the cross-asset spillovers that accompany any habitat shock.

O Equilibrium Characterization

In this section, we describe the characterization of the equilibrium price and quantity processes. We start by considering the case where the model is homoskedastic, so that the conditional covariance matrix is constant. In this case, we can abstract from fluctuations in risk.

O.1 Homoskedastic case

Equilibrium price. In the homoskedastic case, the conditional covariance matrix is constant, so the equilibrium condition can be written as:

$$A_0\, p_t - \sum_{s=1}^{S} A_s\, \mu^p_{t,t+s} = z_t + \sum_{s=1}^{S} B_s\, \mu^d_{t,t+s} + \xi_t \equiv f_t.$$

In the static case, where $A_s = \mathbf{0}_{N\times N}$ for $s \geq 1$, the price is simply $p_t = A_0^{-1} f_t$. To obtain the dynamic version of the equilibrium price, define the conditional expectation operator s periods ahead: $F_t^s x_t \equiv \mathbb{E}_t[x_{t+s}]$. Then, the equilibrium price is given by:

$$p_t = \mathcal{A}(F_t)^{-1} f_t,$$

where $\mathcal{A}(F_t)$ is a polynomial in the forward shift operator F_t,

$$\mathcal{A}(F_t) = A_0 - \sum_{s=1}^{S} A_s\, F_t^s.$$

The inverse of $\mathcal{A}(F)$ is given by:

$$\mathcal{A}(F)^{-1} = \sum_{s=0}^{\infty} \mathcal{A}_s F^s,$$

where the $N \times N$ matrix $\mathcal{A}_s$ satisfies the following recursion:

$$\mathcal{A}_s = \sum_{j=1}^{\min(S,s)} \mathcal{A}_{s-j} A_j A_0^{-1},$$

given the initial condition $\mathcal{A}_0 = A_0^{-1}$.

We can then write the equilibrium price as:

$$p_t = \sum_{s=0}^{\infty} \mathcal{A}_s \mathbb{E}_t[f_{t+s}].$$

We assume that $f_t = H_f X_t$, where $X_t \in \mathbb{R}^K$ follows a VAR(1) process, i.e.,

$$X_t = \Phi X_{t-1} + \nu_t,$$

where $\epsilon_t \sim \mathcal{N}(\mathbf{0}_K, \Sigma_\nu)$.

In this case, we can write the price process as follows:

$$p_t = H_p \Phi^s X_t,$$

where

$$H_p = \sum_{s=0}^{\infty} \mathcal{A}_s H_f \Phi^s.$$

Outside-demand process. We assume that the econometrician observes a subset of the vector of shocks, $\epsilon_{z,t} \in \mathbb{R}^N$, corresponding to the shocks that affect the outside-demand process.

The projection of the outside-demand, dividend, and latent demand processes on the shocks

observed by the econometrician is given by:

$$z_t = \overline{z} + \sum_{h=0}^{\infty} \Psi_h^z \epsilon_{t-h} + \nu_{z,t}$$
$$d_t = \overline{d} + \sum_{h=0}^{\infty} \Psi_h^d \epsilon_{t-h} + \nu_{d,t}$$
$$\xi_t = \overline{\xi} + \sum_{h=0}^{\infty} \Psi_h^\xi \epsilon_{t-h} + \nu_{\xi,t}.$$

We impose that the identified shocks are orthogonal to the latent demand process, i.e., $\mathbb{E}[\xi_t \epsilon_{t-h}] = 0$ for all $h \geq 0$.

O.2 Recovery in the static case

Consider the case of a static demand system, i.e., A_s, B_s, and C_s are zero for $s \geq 1$. In this case, the demand system can be written as:

$$q_t = -A_0 \, p_t + \xi_t.$$

The market clearing condition is given by:

$$q_t = -z_t.$$

Combining the two equations and solving for the equilibrium price, we get:

$$p_t = A_0^{-1}(z_t + \xi_t).$$