# On the Recovery of Demand Elasticities in Dynamic Settings*

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

April 16, 2026

**Abstract**

We study how to recover and use demand elasticities in dynamic asset markets. In demand-system asset pricing, any price movement is accompanied by changes in future expected prices, returns, and risk, so there is no single context-free elasticity. The instrumented elasticity, which captures the reduced-form response to a shock that moves prices together with the expected path of future returns, is shock-specific and depends on the attributes of the identifying variation. The structural elasticity, which holds all future objects fixed, is transportable across shock environments. We show that the inverse of the instrumented elasticity is the price multiplier and characterize how multipliers vary with shock persistence, volatility, systematic exposure, and surprise content. We then develop a recovery theory that uses the impulse responses of flows and prices to extract the transportable structural demand coefficients from shock-specific reduced-form estimates. In weekly data, recovered structural elasticities of roughly 8 to 10 far exceed the naive inverse-multiplier benchmark of about 0.4.

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

**JEL Classification**: G11, G12, G14.

*We thank Ehsan Azarmsa, Aditya Chaudhry, William Fuchs, Zhaoxin Shi, and seminar participants at Boston College and OSU 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

Demand system asset pricing (DSAP) has become a central framework for studying how investor flows shape asset prices and for conducting counterfactual analysis. By modeling investor demand as a function of prices, DSAP has been used to study settings such as ESG (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). The framework is attractive because, once estimated, it can map changes in investor demand into equilibrium price effects and counterfactual repricings. However, DSAP methods (see, among others, Gabaix and Koijen, 2022; Koijen et al., 2024) have focused on a particular type of elasticity: the response to a shock that moves prices together with the entire expected path of future prices and returns. In dynamic asset markets, that is a natural object to study, but it is not the only one.

This multiplicity is not a pathology; it is built into the accounting. The Campbell-Shiller identity requires that any price movement today be absorbed by some combination of higher expected future dividends, higher expected returns at various horizons, or both. A price change therefore cannot be studied in isolation from the path of valuation-relevant objects that move with it. Because there are many ways for a price to move, there is no single demand elasticity in dynamic asset markets: different price processes imply different elasticities even under the same underlying demand system. A low measured elasticity need not mean that investors are intrinsically unwilling to buy when prices fall. It may instead reflect that the identifying shock is persistent and therefore changes not just the current price, but the entire expected path of future prices, returns, and risk.

This observation raises a practical question for counterfactual analysis: what should a researcher do when the counterfactual of interest involves a shock with different persistence, risk profile, or systematic exposure than the shock used for identification? To answer this, we distinguish two

objects. The *structural* demand elasticity is the response of holdings to current prices, holding fixed all future expected prices, dividends, and risk. The *instrumented* demand elasticity is the reduced-form response of holdings to a shock that moves prices together with the expected path of future prices and returns. The instrumented elasticity is shock-specific: it depends on the attributes of the identifying variation. The structural elasticity is not. The instrumented elasticity is the natural object for many DSAP applications, because most are interested in the price impact of persistent or equilibrium flows. The structural elasticity is the transportable primitive that allows a researcher to move from one shock environment to another.

Recent work has highlighted why this distinction matters. Binsbergen, David, and Opp (2025) show that instruments used in the original Koijen and Yogo (2019) and Koijen et al. (2024) frameworks need not recover the structural elasticity when the instrumented price variation persists beyond a single trading period, and that the resulting elasticity estimates may be severely biased or even wrong-signed. Their analysis demonstrates that what appears to be near-zero demand elasticity can be entirely an artifact of shock persistence rather than evidence of structural frictions. Separately, He, Kondor, and Li (2025) show that exogenous supply shifts alter endogenous risk (return volatility, covariance, and the comovement of returns with future investment opportunities), so that the "observed slope" of demand systematically understates the "dynamic slope" derived from investors' first-order conditions by a factor of two to seven.[1]

Our framework speaks to both concerns. By distinguishing structural from instrumented elasticities and showing that the instrumented elasticity is a shock-specific weighted combination of structural demand coefficients, we formalize the sense in which persistence, risk, and other shock attributes enter the measured object. By developing a recovery theory that extracts the structural

[1] Fuchs, Fukuda, and Neuhann (2025) establish a related challenge in a static setting: no-arbitrage cross-asset price linkages make non-parametric identification of demand from supply shocks generically impossible. Our framework is inherently dynamic and sidesteps their impossibility result by working within a parametric linearized demand system and exploiting the time path of impulse responses for identification.

demand system from the path of impulse responses, we show how to move beyond the shock-specific reduced-form elasticity to a transportable structural object.

Our first contribution concerns the instrumented elasticity and its connection to price multipliers. We show that, under linearized market clearing, the inverse of the instrumented elasticity is the price multiplier for the identifying shock. We then show how the instrumented elasticity can be decomposed into structural demand coefficients and shock-specific pass-through terms. When richer impulse responses are observed, our recovery theory goes further and uncovers the underlying structural parameters.

Our second contribution characterizes which shock attributes matter for the instrumented elasticity beyond persistence. In a benchmark CARA/mean-variance equilibrium, we derive four comparative statics: price multipliers are larger when risk is higher, flows are more persistent, shocks load on more systematic directions, and flows are more surprising. We test all four predictions using two non-cash-flow demand shocks, flow-induced trading (FIT) and order-flow imbalance (OFI), and find empirical support for each. These results show that persistence is not the only dimension along which elasticities are context-dependent.

Our third contribution is a recovery theory for the structural demand system. If the measured elasticity depends on the shock process, can one still recover the underlying structural demand? We show that the answer is yes, and that persistent shocks are in fact useful for this purpose: they generate the dynamic impulse responses needed to identify forward-looking demand coefficients. A single contemporaneous multiplier is generally insufficient, but the full path of flow and price responses identifies the structural demand matrices through a convolution equation that links them to the reduced-form impulse responses. We implement this recovery using low-dimensional parameterizations suited to weekly data.

The empirical recovery results show that this distinction is quantitatively large. In weekly

data, the naive inverse-multiplier benchmark, which ignores all forward-looking terms and treats the contemporaneous price response as sufficient, implies an own-price elasticity of about 0.4. Once we use the dynamic path of price and flow responses to recover structural demand, the implied own-price elasticity rises to roughly 8 to 10 in the main combined-shock specification. The reduced-form elasticity is low because the identifying shock is persistent; the structural elasticity is much larger because investors are far more responsive to a price decline expected to reverse than to a persistent repricing that changes the entire path of future expected returns and risk. What looks like weak demand in reduced form is therefore consistent with much stronger structural demand. The recovery result demonstrates that reduced-form impulse responses, while reflecting a shock-specific elasticity, contain enough dynamic information to extract the transportable structural primitives needed for counterfactual analysis in different shock environments.

The rest of the paper is organized as follows. Section 2 sets up the model, including a general log-linearized demand system that nests canonical asset pricing models, and defines equilibrium. Section 3 distinguishes structural from instrumented elasticities and links the latter to the price multiplier. Section 4 imposes a tractable two-parameter demand specification, derives the equilibrium multiplier in closed form, and characterizes how it varies with return variance, flow direction, persistence, surprise content, and investor horizon. Section 5 tests these comparative statics empirically. Section 6 develops the recovery theory and reports the empirical recovery results. Section 7 concludes.

# 2 Model

This section sets up the economic environment. We describe the assets, the investor's log-linearized demand system, and market clearing. We then define equilibrium and discuss the role

of history dependence. The demand system is deliberately general: by restricting its coefficient matrices, it nests the canonical models in the asset pricing literature, including static demand, mean-variance, intertemporal hedging, and adjustment costs. We close by discussing the equivalence between sequential and recursive equilibrium representations and why we adopt the sequential formulation.

## 2.1 Environment

**Assets.** Consider a discrete-time economy with $N$ risky assets indexed by $j = 1, \ldots, N$ and a risk-free asset with 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 $t$, with $p_t \equiv \log P_t$ and $d_t \equiv \log \mathcal{D}_t$. Let $r_{t+1} \in \mathbb{R}^N$ denote the vector of one-period log returns.

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

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

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

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

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-moment state variables are vectors of $m$ unique covariance terms rather than redundant $N \times N$ matrices.

The investor's demand depends on current prices, expected future prices and dividends, and conditional second moments of returns. We write the log-linearized demand system as

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

where $A_0$,$A_s \in \mathbb{R}^{N\times N}$, $B_s \in \mathbb{R}^{N\times N}$, and $D_s \in \mathbb{R}^{N\times m}$. The matrix $A_0$ governs the contemporaneous response of holdings to current prices. The matrices $A_s$ and $B_s$ capture how expectations of future prices and dividends at horizon $s$ influence current demand. The matrices $D_s$ capture the effect of expected future return risk; under this sign convention, larger $D_s$ means riskier future states depress demand more strongly. The vector $\xi_t$ is a latent demand shifter, unobserved and potentially correlated with equilibrium prices, in the sense of Koijen and Yogo (2019). This formulation is general enough to accommodate a wide range of demand specifications, from static models to dynamic models with frictions, as we show in Section 2.3.

**Campbell–Shiller accounting identity.** The one-period Campbell–Shiller approximation relates log returns to log prices and log dividends:

$$r_{t+1} = \kappa + \rho\, p_{t+1} + (1-\rho)\, d_{t+1} - p_t, \tag{2}$$

where $\rho \in (0,1)$ and $\kappa$ is a log-linearization constant. Iterating forward under a transversality condition yields the long-run representation

$$p_t = \bar{\kappa} + (1-\rho) \sum_{j=1}^{\infty} \rho^{j-1}\, \mathbb{E}_t[d_{t+j}] - \sum_{j=1}^{\infty} \rho^{j-1}\, \mathbb{E}_t[r_{t+j}], \tag{3}$$

where $\bar{\kappa}$ collects constants. Differentiating (3) with respect to $p_t$ gives the accounting restriction

$$I = (1-\rho)\sum_{j=1}^{\infty}\rho^{j-1}\frac{\partial \mathbb{E}_t[d_{t+j}]}{\partial p_t^\top} - \sum_{j=1}^{\infty}\rho^{j-1}\frac{\partial \mathbb{E}_t[r_{t+j}]}{\partial p_t^\top}. \tag{4}$$

Equation (4) is the key constraint. A price movement cannot occur in isolation: it must pass through to expected dividends, expected returns, or both. Any elasticity concept therefore has to specify that pass-through pattern. Different assumptions about what adjusts when prices move produce different elasticities, even under the same underlying demand system.

**Return-based demand and coefficient restrictions.** To make this multiplicity of relevant elasticities precise, it is useful to write the demand function in return form. Using the Campbell–Shiller identity (2) to substitute expected returns for expected prices and dividends restricts the coefficient matrices and yields

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

where $C_s \in \mathbb{R}^{N\times N}$ describes how current holdings respond to expected returns at horizon $s$. The price–dividend coefficients $(A_0, A_s, B_s)$ and the return coefficients $C_s$ are linked by the Campbell–Shiller restrictions:

$$A_0 = C_1, \qquad A_s = \rho\,C_s - C_{s+1} \;\text{ for } s = 1,\ldots,S, \qquad B_s = (1-\rho)\,C_s \;\text{ for } s = 1,\ldots,S, \tag{5}$$

with the convention $C_{S+1} \equiv 0$. In this sense, the $C_s$ matrices are primitive and the $A_s$, $B_s$ matrices are derived objects.

**From demand coefficients to preferences.** The coefficient matrices $(A_0, A_s, B_s, D_s)$ are not free parameters. When the demand system (1) is generated by an investor who maximizes an expected utility objective, the coefficients inherit restrictions from the underlying preference ordering. By the integrability theorem in classical demand theory,[2] these restrictions are also sufficient: knowledge of the demand coefficients is enough to recover the preference parameters that generated them, up to monotone transformations of the utility function. This provides a formal justification for treating the structural demand coefficients as the primitive objects of interest: they encode all the preference information that is relevant for counterfactual analysis.

**Reduced-form representation.** Because our main results concern the price channel of demand, it is convenient to isolate the dependence on current and expected future prices by absorbing conditional second moments, expected dividends, and demand shocks into a reduced-form shifter:

$$q_t = -A_0\, p_t + \sum_{s=1}^{S} A_s\, \mu^p_{t,t+s} + f_t, \tag{6}$$

where

$$f_t \equiv \sum_{s=1}^{S} B_s\, \mu^d_{t,t+s} - \sum_{s=1}^{S} D_s\, v^r_{t,t+s} + \xi_t.$$

**Market clearing.** Let $Z_t \in \mathbb{R}^N$ denote outside (inelastic) holdings, with reference point $\bar{Z}$. Total shares outstanding are fixed and normalized so that

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

[2] See Mas-Colell, Whinston, and Green (1995, Chapter 3). Appendix C states the theorem and discusses its implications for the demand coefficients.

Linearizing around $(\bar{Q},\bar{Z})$ with $\bar{Q} + \bar{Z} = \mathbf{1}_N$ yields

$$\mathrm{Diag}(\bar{Q})\, q_t + z_t = 0, \qquad z_t \equiv Z_t - \bar{Z}.$$

Define $\Lambda \equiv \mathrm{Diag}(\bar{Q})^{-1}$. Then market clearing implies

$$q_t = -\Lambda\, z_t. \tag{7}$$

Substituting (7) into (6) gives the linear equilibrium condition

$$A_0\, p_t - \sum_{s=1}^{S} A_s\, \mu^p_{t,t+s} = \Lambda\, z_t + f_t. \tag{8}$$

This forward-looking equation links current prices to the expected path of future prices. Its solution, and thus the equilibrium price process, depends on the stochastic environment, in particular on the history of all shocks that drive the forcing terms on the right-hand side.

## 2.2 Equilibrium and history dependence

**Structural shocks.** Let $\nu_t = (\nu_{t,1}, \ldots, \nu_{t,K})$ denote a $K$-dimensional vector of structural 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 $\nu_t$ collects all primitive sources of randomness in the economy. In the context of the demand system (1), it includes innovations to outside holdings $z_t$, to the latent demand shifter $\xi_t$, and to whatever processes govern the expected paths of future dividends $\{\mu^d_{t,t+s}\}$ and return volatility $\{v^r_{t,t+s}\}$.

**History dependence of equilibrium objects.** In equilibrium, prices and dividends are functions of the entire history of shocks:

$$p_t = p(\nu^t, t), \qquad d_t = d(\nu^t, t).$$

The same is true for all objects that enter the demand system. Conditional expectations and conditional second moments are determined by the filtration generated by $\nu^t$:

$$\mu^p_{t,t+s} = \mu^p(\nu^t, t, s), \qquad v^r_{t,t+s} = v^r(\nu^t, t, s).$$

Writing equilibrium objects as functions of the *full* history $\nu^t$ is deliberately general: we do not assume that a simpler set of state variables summarizes all that is relevant about the past. In particular, even if the structural shocks are i.i.d., the equilibrium price can inherit persistence through the demand system's dependence on expected future prices.

Different structural shocks will generally have different effects on these equilibrium objects. A shock that is transitory may move current prices without affecting expectations, while a persistent shock may shift the entire expected path of future prices, returns, and risk. Because the demand system loads on all these objects, the total response of demand to a shock (and hence the reduced-form elasticity) is shock-specific even though the underlying structural demand coefficients are invariant.

**Equilibrium definition.**

**Definition 1** (Equilibrium)**.** *Given a stochastic process for outside holdings $\{z_t\}$ and a demand shifter process $\{f_t\}$, an equilibrium is a pair of stochastic processes $\{p_t, q_t\}$ adapted to the filtration generated by $\nu^t$ 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* $Diag(\bar{Q})\, q_t + z_t = 0$ *holds at every date* $t$.

Because demand depends on expected future prices, the equilibrium condition (8) is a forward-looking difference equation: current prices appear on the left-hand side through $A_0\, p_t$, while expected future prices $\mathbb{E}_t[p_{t+s}]$ appear through the coefficients $A_s$. Solving this equation requires ruling out explosive (bubble) paths. Under a standard transversality condition, the unique bounded solution expresses the current price as a discounted sum of current and expected future values of the forcing term $\Lambda\, z_t + f_t$. Intuitively, the price today reflects not only today's demand imbalance but also the market's expectation of all future imbalances, weighted by how sensitive demand is to prices at each horizon. When the forcing term depends on the full history of shocks $\nu^t$, the equilibrium price inherits that dependence.

## 2.3 Nesting canonical models

The demand system (1) is deliberately general. By restricting the coefficient matrices $(A_0, A_s, B_s, D_s)$, it nests the main demand specifications used in the asset pricing and DSAP literatures.

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

$$q_t = -A_0\, p_t + f_t.$$

This is the specification that underlies much of the initial DSAP literature, including Koijen and Yogo (2019) and Koijen et al. (2024). In this model, investors respond only to current prices; the price elasticity is $A_0$, and there are no intertemporal links. Because demand ignores future

expectations, any instrument that shifts current prices recovers the same elasticity $A_0$ regardless of how persistent the shock is.

**Mean–variance (myopic) demand.** Setting $S = 1$ restricts the investor to care only about one-period-ahead prices and dividends:

$$q_t = -A_0\, p_t + A_1\, \mu^p_{t,t+1} + B_1\, \mu^d_{t,t+1} - D_1\, v^r_{t,t+1}. \tag{9}$$

The investor is myopic: only next-period expectations influence current holdings.

A concrete microfoundation is the mean–variance investor with constant risk aversion $\gamma > 0$ and perceived dollar covariance matrix $\Sigma \succ 0$. Exact demand in levels is

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

Log-linearizing around $(\bar{Q},\bar{P},\bar{\mathcal{D}})$ delivers (9) with

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

where $\beta = 1/R^f$. The relation $A_1 = \beta\, A_0$ reflects the fact that, in this model, the investor cares about prices and dividends only through their effect on next-period expected returns.

When second moments are constant, $D_1$ drops out and the model is a pure expected-return demand system. When second moments are time-varying, the $-D_1\, v^r_{t,t+1}$ term captures the investor's sensitivity to changes in risk.

**Intertemporal hedging demand.** Merton (1973) shows that a long-horizon investor with CRRA preferences holds not only the myopic mean–variance portfolio but also an additional *hedging*

portfolio that insures against adverse shifts in future investment opportunities. The hedging component makes demand sensitive to expected prices and dividends at horizons $s \geq 2$, so the demand system requires $S > 1$ with coefficient matrices $A_s$ and $B_s$ whose values are determined by the investor's risk aversion, horizon, and the dynamics of the investment opportunity set.

The log-linear approximation of Campbell and Viceira (2002) is written in terms of portfolio *weights* $\alpha_t$ (fractions of wealth), not quantities, so mapping it into our demand system requires an additional step. The optimal weight vector is linear in the state variables that predict future returns, with a myopic component proportional to expected excess returns and a hedging component that depends on the covariance of returns with changes in investment opportunities. To map from weights to the log quantities in our demand system, note that the number of shares held is $Q_{i,t} = \alpha_{i,t} W_t / P_{i,t}$, so in logs

$$q_{i,t} = \tilde{\alpha}_{i,t} + w_t - p_{i,t},$$

where $\tilde{\alpha}_{i,t}$ is the log-deviation of the portfolio weight and $w_t \equiv \log W_t - \log \bar{W}$ is the log-deviation of wealth. Because the Campbell and Viceira (2002) portfolio weights are linear in state variables, the demand in log quantities inherits the same linear structure. The coefficient matrices $A_s$ and $B_s$ below therefore incorporate both the Campbell and Viceira hedging coefficients and the steady-state wealth-to-price ratio $\bar{W}/\bar{P}_i$, while the endogenous wealth term $w_t$ is absorbed into the demand shifter $\xi_t$.[3]

[3]Unlike the CARA case above, CRRA demand depends on wealth. The wealth term $w_t$ is itself a function of past returns and consumption, hence of the history $v^t$. Our sequential formulation handles this naturally: $w_t$ enters $\xi_t$ as part of the full-history dependence, without requiring us to specify how wealth evolves as a function of a particular state vector.

In our general framework, the hedging component appears as

$$q_t = \underbrace{-A_0\, p_t + A_1\, \mu^p_{t,t+1} + B_1\, \mu^d_{t,t+1}}_{\text{myopic component}} + \underbrace{\sum_{s=2}^{S} A_s\, \mu^p_{t,t+s} + \sum_{s=2}^{S} B_s\, \mu^d_{t,t+s}}_{\text{hedging component}} - \sum_{s=1}^{S} D_s\, v^r_{t,t+s}.$$

The key distinction from the mean–variance case is that the hedging coefficients $A_s$ and $B_s$ for $s \geq 2$ need not be proportional to $A_1$ and $B_1$ or decay at any particular rate. Their magnitude and sign depend on the risk aversion of the investor, the persistence of the state variables that drive expected returns, and the covariance structure of returns with those state variables.

The presence of hedging demand is economically important for the elasticity analysis. Because the hedging component responds to longer-horizon expectations, a persistent shock that shifts the entire expected path of future prices will elicit a different demand response than a transitory shock that moves only next-period expectations. This is one of the channels through which reduced-form elasticities become shock-specific even under the same structural demand system.

**Adjustment costs.** Gârleanu and Pedersen (2013) study the optimal trading strategy of a mean–variance investor who faces quadratic transaction costs. The key insight is that the optimal portfolio is not the current frictionless optimum but rather an "aim" portfolio, a forward-looking weighted average of future frictionless optimal portfolios, toward which the investor trades partially each period.

In the Gârleanu and Pedersen (2013) model, the investor's position $Q_t$ evolves according to

$$Q_t - Q_{t-1} = \lambda_{\text{tc}}\big(\text{Aim}_t - Q_{t-1}\big), \tag{10}$$

where $\lambda_{\text{tc}} \in (0,1)$ governs the speed of adjustment to the aim portfolio. The aim portfolio itself is

$$\text{Aim}_t = \sum_{s=0}^{\infty} \zeta^s \, Q^*_{t+s|t},$$

where $Q^*_{t+s|t}$ denotes the expected frictionless optimum at date $t + s$ given time-$t$ information and $\zeta \in (0,1)$ depends on the transaction cost parameter.

To see how this maps into our framework, note that the frictionless optimum $Q^*_{t+s|t}$ is itself a function of expected future prices and dividends at date $t + s$ (the mean–variance demand). The aim portfolio therefore loads on the entire sequence of expected future prices and dividends with geometrically declining weights:

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

In the log-linear demand notation, the geometric weighting of future frictionless optima generates price–dividend coefficients that decay geometrically:

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

where $A_1$ and $B_1$ are the myopic (frictionless) coefficients and $\zeta \in (0,1)$ encodes the rate at which the investor discounts future frictionless targets. This geometric-decay parameterization is the two-parameter generalization studied empirically in the recovery analysis.

Although the underlying preference is myopic, transaction costs make demand dynamic: $A_s \neq 0$ and $B_s \neq 0$ for all $s \geq 1$, because adjusting today is costly and the investor optimally anticipates future desired positions. The decay rate $\zeta$ is pinned down by the transaction cost parameter, with higher costs implying slower adjustment and a longer effective horizon. The partial-adjustment

| Model | Horizon $S$ | Coefficient restrictions | References |
|---|---|---|---|
| Static | 0 | $A_s = B_s = 0$ for all $s \geq 1$ | Koijen and Yogo (2019), Gabaix and Koijen (2022) |
| Mean–variance | 1 | $A_1 = \beta A_0$; $A_s = B_s = 0$ for $s \geq 2$ | Mean–variance CAPM |
| Hedging demand | $S > 1$ | $A_s, B_s$ unrestricted | Merton (1973), Campbell and Viceira (2002) |
| Adjustment costs | $\infty$ | $A_s \propto \zeta^{s-1} A_1$; $B_s \propto \zeta^{s-1} B_1$ | Gârleanu and Pedersen (2013) |

**Table 1. Nesting canonical demand models.**
Each row shows the restrictions on the demand coefficients $\{A_s, B_s\}_{s=1}^{S}$ that reduce the general demand system (1) to a canonical special case. The static model ignores all future expectations. The mean–variance model responds only to one-period-ahead expected prices and dividends. Merton's hedging demand adds unrestricted sensitivity to longer-horizon expectations. The Gârleanu–Pedersen model generates geometrically decaying coefficients through the aim-portfolio structure induced by quadratic adjustment costs.

structure in (10) also makes the current portfolio depend on the lagged portfolio $Q_{t-1}$, introducing state dependence that is absorbed into the demand shifter $f_t$ in our reduced-form representation.

Table 1 summarizes how the coefficient restrictions define each canonical model.

## 2.4 Sequential and recursive equilibrium representations

The demand system (1) is cast in sequential form: at each date $t$ the investor conditions on the full history $v^t$ and forms expectations about the entire future path of prices, dividends, and risk. In the terminology of Ljungqvist and Sargent (2018, Chapter 8), agents trade period by period in spot markets, and all equilibrium objects are functions of the history $v^t$ with no restriction on how the past enters.[4] A recursive competitive equilibrium (RCE) imposes additional structure: it requires that all equilibrium objects depend on $v^t$ only through a finite-dimensional Markov state vector $s_t$. Our sequential formulation nests any such RCE as a special case, because any function of $s_t$ is also a function of the history $v^t$ that determines $s_t$. The converse does not hold: the sequential form accommodates settings where no finite-dimensional state suffices. In particular, the demand

[4] Appendix D provides a detailed discussion.

coefficients $\{A_s, B_s, D_s\}$ are defined without taking a stand on which state variables generate the conditional expectations $\{\mu^p_{t,t+s}, \mu^d_{t,t+s}, v^r_{t,t+s}\}$.

This generality has two concrete payoffs. First, the structural coefficients $\{A_s, B_s, D_s\}$ are properties of the investor's demand function, not of a particular equilibrium model or shock process. They are therefore transportable across shock environments. Second, the equilibrium condition (8) links the structural demand coefficients $\{A_s\}$ to equilibrium impulse responses through a convolution equation (derived in Section 6):

$$A_0 \, \Xi_k - \sum_{s=1}^{S} A_s \, \Xi_{k+s} = \Psi_{y,k}, \qquad k \geq 0, \tag{11}$$

where $\Xi_k$ denotes the reduced-form price impulse response at horizon $k$ to an identified demand shock (i.e. the $k$-period-ahead response of log prices to a unit innovation in outside demand), and $\Psi_{y,k}$ is the corresponding impulse response of the composite demand shifter $y_t \equiv \Lambda z_t + f_t$. Because the sequential form keeps the entire expected path as arguments, the impulse responses of prices and flows can be used to recover the structural demand matrices without assuming a law of motion for the state.

To illustrate the contrast, Binsbergen et al. (2025) study demand elasticities in a dynamic setting using the recursive approach: they specify a continuous-time CRRA general-equilibrium model with an explicit Markov state vector (the "shifter state" driving exogenous price shifts), solve for value functions and optimal portfolio policies via the Hamilton–Jacobi–Bellman (HJB) equation, and characterize equilibrium as a fixed point. The recursive approach yields a complete structural model suited to counterfactual analysis, but it requires the researcher to commit to a particular set of state variables and functional forms. Our sequential formulation nests any such recursive model as a special case while remaining agnostic about the choice of state vector.

# 3 Structural vs. Instrumented Elasticities

In dynamic asset markets, any price movement is accompanied by changes in future expected prices, returns, and risk, so there is no single context-free demand elasticity. The Campbell–Shiller identity requires that a price change today be absorbed by some combination of higher expected future dividends, higher expected returns at various horizons, or both. Because there are many possible configurations of how expectations of future variables accompany the current price movement, the measured elasticity depends on which configuration the identifying variation generates, even under the same underlying demand system. A low measured elasticity need not mean that investors are intrinsically unwilling to buy when prices fall; it may instead reflect that the identifying shock is persistent and therefore changes not just the current price, but the entire expected path of future returns and risk.

This section draws a sharp line between two objects. The *structural* demand elasticity is the response of holdings to current prices, holding fixed all future expected prices, dividends, and risk. The *instrumented* demand elasticity is the reduced-form response to a shock that moves prices together with the expected path of future prices, returns, and risk. The instrumented elasticity is shock-specific: it depends on the persistence, risk profile, and systematic exposure of the identifying variation. The structural elasticity is not; it is the transportable primitive that allows a researcher to move from one counterfactual scenario to another. We show that the instrumented elasticity equals the inverse of the price multiplier. We then decompose it into the structural demand coefficients and characterize the special cases that arise when the identifying shock operates through different channels.

## 3.1 Structural demand elasticities

**Contemporaneous structural elasticity.** Recall the reduced-form demand system (6). Holding fixed all future expected prices $\mu^p_{t,t+s}$ and the composite shifter $f_t$, the local elasticity of log holdings with respect to current log prices is

$$-\frac{\partial q_t}{\partial p_t^\top} = A_0.$$

This is a partial-equilibrium object: the slope of demand with respect to current prices when every other argument of the demand system is held constant.

At first glance, $A_0$ might seem like the only "true" demand elasticity, since it is the derivative of current holdings with respect to current prices holding all future objects fixed. But that interpretation is too narrow for dynamic asset markets. The Campbell–Shiller identity (introduced in detail in Section 3.2 below) imposes an accounting constraint: if current prices move, then some combination of expected future returns and expected future dividends must move as well for beliefs to remain internally consistent. Once that consistency requirement is imposed, many consistent structural elasticities become relevant, each corresponding to a different assumption about how a price movement is absorbed.

## 3.2 Instrumented elasticity and the price multiplier

**Instrumented elasticity.** The instrumented elasticity (IE) matrix is the ratio of the reduced-form demand response to the first-stage price response:

$$\mathcal{E}_t(z_t) \equiv \underbrace{-\frac{dq_t}{dz_t^\top}}_{\text{reduced form}} \left( \overbrace{\frac{dp_t}{dz_t^\top}}^{\text{first stage}} \right)^{-1}. \tag{12}$$

This is the population analog of an instrumental-variables estimator: demand regressed on prices, instrumented by the demand shock $z_t$.

**Price multiplier.** The first-stage matrix in (12) defines the price multiplier:

$$\mathcal{M}_t(z_t) \equiv \frac{dp_t}{dz_t^\top}.$$

## 3.3 Demand shock exogeneity and the main equivalence

**Assumption 1** (Latent demand exogeneity)**.** *Let the identifying outside-demand process admit the moving-average representation*

$$z_t = \bar{z} + \sum_{h=0}^{\infty} \Psi_h^z \varepsilon_{t-h},$$

*where $\varepsilon_t$ is the identified exogenous innovation in outside demand at date $t$, and $\Psi_h^z$ is the horizon-$h$ impulse response of $z_{t+h}$ to $\varepsilon_t$. The latent demand shifter is exogenous to both the current outside-demand state $z_t$ and the identified innovations $\{\varepsilon_t\}$:*

$$\frac{d\xi_t}{dz_t^\top} = 0.$$

*Moreover, in the linear projection of $\xi_t$ on the shock history $\{\varepsilon_{t-h}\}_{h\geq 0}$,*

$$\xi_t = \bar{\xi} + \sum_{h=0}^{\infty} \Psi_h^\xi \varepsilon_{t-h}, \qquad \Psi_h^\xi = 0 \quad \textit{for all } h \geq 0.$$

This assumption requires that the identifying shock does not directly shift the unobserved component of demand. In the language of Koijen and Yogo (2019), it rules out a correlation

between the instrument and the latent demand for unobserved characteristics. Assumption 1 ensures that the instrument affects demand only through the equilibrium price path.

**Equivalence result.** We provide our equivalence result below in Proposition 1.

**Proposition 1** (Instrumented elasticity and the multiplier)**.** *Suppose the price multiplier $\mathcal{M}_t(z_t)$ is invertible, linearized market clearing* (7) *holds, and Assumption 1 holds. Then*

$$\mathcal{E}_t(z_t) = \Lambda\, \mathcal{M}_t(z_t)^{-1}. \tag{13}$$

*Define the aggregate elasticity $\tilde{\mathcal{E}}_t(z_t) \equiv \Lambda^{-1}\mathcal{E}_t(z_t) = Diag(\bar{Q})\, \mathcal{E}_t(z_t)$. Then*

$$\tilde{\mathcal{E}}_t(z_t) = \mathcal{M}_t(z_t)^{-1}. \tag{14}$$

See Appendix A for the proof. The aggregate elasticity $\tilde{\mathcal{E}}_t(z_t)$ rescales the investor-level IE into shares-outstanding units, matching the units of $z_t$. In those units, the aggregate elasticity is the inverse of the price multiplier.[5]

## 3.4 What the instrumented elasticity loads on

The instrumented demand elasticity is a reduced form quantity. It is a weighted combination of the structural demand coefficients, with weights determined by how the identifying shock passes through to future expected prices, dividends, and risk. For each horizon $s$, define the shock-specific

[5]With multiple elastic investors, the aggregate IE is the holdings-weighted sum of their instrumented elasticities in shares-outstanding units. The outside investor contributes zero because its demand is inelastic. See Section 6.3 below for the extension with heterogeneous investors.

pass-through matrices

$$\Pi^{p}_{t,s}(z_t) \equiv \frac{d\mu^{p}_{t,t+s}}{dz_t^\top}\left(\frac{dp_t}{dz_t^\top}\right)^{-1}, \qquad \Pi^{d}_{t,s}(z_t) \equiv \frac{d\mu^{d}_{t,t+s}}{dz_t^\top}\left(\frac{dp_t}{dz_t^\top}\right)^{-1}, \qquad \Pi^{v}_{t,s}(z_t) \equiv \frac{dv^{r}_{t,t+s}}{dz_t^\top}\left(\frac{dp_t}{dz_t^\top}\right)^{-1}.$$

These objects are the characteristic analogs of the instrumented elasticity itself: each is the reduced-form response of the corresponding characteristic to the shock, divided by the first-stage price response.

**Proposition 2** (Instrumented elasticity and shock-specific pass-through)**.** *Suppose the price multiplier $\mathcal{M}_t(z_t) = dp_t/dz_t^\top$ is invertible and Assumption 1 holds. Then the instrumented elasticity satisfies*

$$\mathcal{E}_t(z_t) = A_0 - \sum_{s=1}^{S} A_s\,\Pi^{p}_{t,s}(z_t) - \sum_{s=1}^{S} B_s\,\Pi^{d}_{t,s}(z_t) + \sum_{s=1}^{S} D_s\,\Pi^{v}_{t,s}(z_t), \tag{15}$$

See Appendix A for the proof.

The IE therefore loads on three channels: pass-through of the price movement to future expected prices, to future expected dividends, and to future return risk. The structural elasticity $A_0$ is the baseline; the remaining terms capture how the instrument propagates through each characteristic at each horizon.

## 3.5 Special cases

Equation (15) shows that the IE generically depends on all the structural demand coefficients and all the pass-through matrices induced by the instrument. Only under stronger restrictions on the shock does the IE isolate narrower structural objects.

**Static IE.** If the instrument induces a purely transitory price change with no pass-through to future expected prices, dividends, or risk ($\Pi^p_{t,s} = \Pi^d_{t,s} = \Pi^v_{t,s} = 0$ for all $s$), then

$$\mathcal{E}^{\text{stat}}_t(z_t) = A_0.$$

The IE coincides with the structural elasticity. This is the only case in which the instrument identifies $A_0$ directly.

**Variance IE.** If the instrument changes conditional second moments while leaving expected future prices and dividends unchanged ($\Pi^p_{t,s} = \Pi^d_{t,s} = 0$ for all $s$), then

$$\mathcal{E}^{\text{var}}_t(z_t) = A_0 + \sum_{s=1}^{S} D_s \, \Pi^v_{t,s}(z_t).$$

This covers shocks that operate through future return variance, dividend variance, or price-dividend covariance.

**Immediate IE.** If the instrument generates a pure one-period price movement that passes through entirely to expected prices one period ahead and nothing else ($\Pi^p_{t,1} = I$, $\Pi^p_{t,s} = 0$ for $s \geq 2$, $\Pi^d_{t,s} = \Pi^v_{t,s} = 0$), then

$$\mathcal{E}^{\text{imm}}_t(z_t) = A_0 - A_1.$$

The IE picks up the contemporaneous price sensitivity net of the one-period-ahead expected price sensitivity. Under myopic mean-variance demand with $A_1 = \beta \, A_0$, this reduces to $(1 - \beta) \, A_0$.

**Dynamic IE.** If the instrument shifts the entire expected path of future prices and possibly future risk, then

$$\mathcal{E}_t^{\text{dyn}}(z_t) = A_0 - \sum_{s=1}^{S} A_s \, \Pi_{t,s}^{p}(z_t) - \sum_{s=1}^{S} B_s \, \Pi_{t,s}^{d}(z_t) + \sum_{s=1}^{S} D_s \, \Pi_{t,s}^{v}(z_t).$$

This is the empirically relevant case. Persistent flow instruments generally identify neither $A_0$ nor any single coefficient matrix; they identify a weighted combination of all the structural demand coefficients, with weights determined by how the shock propagates through expected prices, dividends, and risk at each horizon.

## 3.6 There is no single price elasticity

These special cases make clear why searching for "the" demand elasticity is misguided in dynamic asset markets. Different instruments generate different pass-through patterns across horizons and across the price, dividend, and risk channels, and therefore recover distinct objects even when all are consistent with the same underlying structural demand system. The multiplicity of IEs is not a defect of any particular instrument; it is a consequence of the fact that each IE bundles the structural demand slopes with the propagation of that specific shock through the full set of demand-relevant state variables.

This is why counterfactual analysis requires the structural coefficients, not an IE. Much of the DSAP literature estimates price multipliers using persistent instruments that shift not just current prices but the entire expected path of future prices and returns. Such designs do not recover a ceteris-paribus elasticity that holds expectations fixed. They recover IEs that blend structural demand slopes with the propagation of the instrument through prices, dividends, and risk at each horizon. The practical problem is immediate: when the counterfactual of interest involves a shock with different persistence, risk profile, or systematic exposure than the shock used for identification,

the IE estimated from the data cannot be transported to the new scenario. The structural demand coefficients $\{A_s, B_s, D_s\}$ can, because they are properties of the investor's demand function rather than of any particular shock process.

Asset demand is indirect: investors do not value prices per se, but the return and cash-flow characteristics that prices encode. Once that pass-through is made explicit, the multiplicity of elasticities is no longer puzzling. Each elasticity corresponds to a distinct ceteris-paribus clause about how a price movement reallocates those characteristics across horizons. The challenge, taken up in Section 6, is to recover the transportable structural coefficients from the shock-specific reduced-form evidence.

# 4 Elasticity as a Function of Key Variables

Sections 2 and 3 have emphasized that dynamic environments admit multiple elasticity concepts and that instrumented elasticities generally differ from structural derivatives. This section shows that multipliers and instrumented elasticities are still tractable and economically interpretable in equilibrium. We study a demand system parameterized by two scalars $(\beta,\delta)$ that control the investor's sensitivity to future expected prices, derive the equilibrium multiplier and instrumented elasticity, and then characterize the comparative statics that guide the empirical analysis. The central lesson is constructive: in the presence of persistent flows, the relevant object for mapping flows to prices is a dynamic IE, and its determinants can be characterized sharply.

## 4.1 A tractable demand specification

The general demand system (1) leaves the coefficient matrices $(A_0, A_s, B_s, D_s)$ unrestricted. To connect with many commonly used canonical models in asset pricing and to derive comparative

statics that guide the empirical analysis, we impose the two-parameter geometric-decay structure shown in Section 2.3:

$$A_s = \beta\,\delta^{s-1}\,A_0, \qquad s = 1,2,\ldots, \tag{16}$$

where $\beta \in [0,1)$ and $\delta \in [0,1)$. The parameter $\beta$ controls the overall sensitivity of demand to one-period-ahead expected prices relative to current prices; in the mean–variance microfoundation of Section 2.3, $\beta = 1/R^f$ is the discount factor. The parameter $\delta$ controls the rate at which sensitivity decays across horizons; in the Gârleanu and Pedersen (2013) model, $\delta$ corresponds to the aim-portfolio decay rate $\zeta$ in Section 2.3.

The parameterization nests the canonical special cases. When $\beta = 0$, all intertemporal coefficients vanish ($A_s = 0$ for $s \geq 1$) and demand responds only to current prices (static demand). When $\delta = 0$ and $\beta = 1/R^f$, only the one-period-ahead coefficient survives ($A_1 = \beta\,A_0$, $A_s = 0$ for $s \geq 2$), which is the myopic mean–variance model. When $\delta > 0$ and $\beta = 1/R^f$, the full sequence of coefficients $A_s = \beta\,\delta^{s-1}\,A_0$ is nonzero, capturing the forward-looking demand induced by adjustment costs in the Gârleanu and Pedersen (2013) model.

Under (16), the reduced-form demand (6) becomes

$$q_t = -A_0\,p_t + \beta\,A_0 \sum_{s=1}^{\infty} \delta^{s-1}\,\mathbb{E}_t[p_{t+s}] + f_t.$$

## 4.2 Equilibrium with persistent outside flows

To isolate the price channel, suppose expected dividends and conditional second moments are constant, so that $f_t$ is absorbed into the intercept. Using market clearing (7), the equilibrium condition becomes

$$A_0\,p_t - \beta\,A_0 \sum_{s=1}^{\infty} \delta^{s-1}\,\mathbb{E}_t[p_{t+s}] = \Lambda\,z_t. \tag{17}$$

Assume outside holdings follow

$$z_{t+1} = \phi\, z_t + \varepsilon_{t+1} + \eta_t, \qquad \phi \in [0,1), \tag{18}$$

where $\varepsilon_{t+1}$ is a surprise innovation in outside demand, and $\eta_t$ is an *expected* innovation in outside demand that is *announced* at time $t$ but realized at time $t+1$. This allows us to later discuss differential price multipliers for surprise versus expected demand.

**Proposition 3** (Unique equilibrium under geometric-decay demand)**.** *Suppose* $(\beta+\delta)\phi < 1$. *Then the equilibrium condition* (17) *admits the unique stationary solution*

$$p_t = \mathcal{M}\, z_t + \frac{\beta}{1-\delta\phi}\,\mathcal{M}\,\eta_t, \qquad \mathcal{M} = \frac{1-\delta\phi}{1-(\beta+\delta)\phi}\,(A_0)^{-1}\Lambda. \tag{19}$$

See Appendix A for the proof.

Proposition 3 shows that equilibrium prices are linear in both the realized outside position $z_t$ and the expected innovation $\eta_t$. The coefficient on $\eta_t$ is $\frac{\beta}{1-\delta\phi}\,\mathcal{M}$. When demand is myopic ($\delta = 0$), the coefficient reduces to $\beta\mathcal{M}$, recovering the mean–variance benchmark.

The multiplier $\mathcal{M}$ is the log-price multiplier matrix: it maps percentage-of-shares demand shocks into percentage price responses. Its structure reveals two distinct sources of amplification. The factor $(A_0)^{-1}\Lambda$ is the *static* multiplier, the price response that would obtain if demand were purely contemporaneous. The scalar

$$\alpha(\beta,\delta,\phi) \equiv \frac{1-\delta\phi}{1-(\beta+\delta)\phi} \tag{20}$$

is the *dynamic amplification* generated by the interaction of forward-looking demand and persistent flows.

When demand is static ($\beta = 0$), the investor ignores future expectations, so $\alpha = 1$ and $\mathcal{M} = (A_0)^{-1}\Lambda$ with no dynamic amplification. When demand is myopic ($\delta = 0$), the amplification factor reduces to $\alpha = 1/(1 - \beta\phi)$, recovering the mean–variance benchmark in Appendix B. When flows are transitory ($\phi = 0$), $\alpha = 1$ regardless of $(\beta,\delta)$, because a shock that carries no information about future imbalances cannot trigger any feedback.

## 4.3 Instrumented elasticity

In the stationary environment of this section, the multiplier and IE are time-invariant. Because market clearing gives $dq_t/dz_t^\top = -\Lambda$, the investor-level IE is

$$\mathcal{E} = \Lambda\,\mathcal{M}^{-1} = \frac{1 - (\beta + \delta)\phi}{1 - \delta\phi}\,A_0. \tag{21}$$

The corresponding aggregate elasticity is $\tilde{\mathcal{E}} \equiv \Lambda^{-1}\mathcal{E} = \mathcal{M}^{-1}$. The inverse multiplier is not the structural elasticity $A_0$; it is the aggregate instrumented elasticity, which scales $A_0$ by the reciprocal of the dynamic amplification factor (20):

$$\tilde{\mathcal{E}} = \alpha(\beta,\delta,\phi)^{-1}\,\Lambda^{-1}A_0.$$

When $\delta = 0$, the IE simplifies to $\mathcal{E} = (1 - \beta\phi)\,A_0$, the expression that obtains in the myopic mean–variance benchmark.

## 4.4 What does a flow instrument identify?

Recall from Section 3.5 that different instruments generate different pass-through patterns and therefore identify different objects. When outside holdings are persistent ($\phi > 0$), a shock to $z_t$

moves not only the current price but the entire expected path $\{\mathbb{E}_t[p_{t+s}]\}_{s\geq 1}$, shifting the whole expected-return term structure rather than only next-period expected returns.

This is why the object in (21) should be interpreted as an investor-level dynamic IE. Persistent flow instruments do not identify the ceteris-paribus derivative that holds fixed all future expectations. They identify the elasticity relevant for the equilibrium mapping from persistent flows to prices.

In the $(\beta,\delta)$ model that distinction is especially transparent. The structural elasticity is $A_0$. The investor-level IE is $\frac{1-(\beta+\delta)\phi}{1-\delta\phi} A_0$, which is strictly smaller than $A_0$ whenever $\beta\phi > 0$. The gap between the two grows with flow persistence $\phi$, with the one-period-ahead sensitivity $\beta$, and with the forward-looking horizon $\delta$. Each source of dynamic feedback pushes the IE further from the structural derivative.

## 4.5 Multiplier comparative statics

The general multiplier (19) delivers comparative statics with respect to each of its determinants. All five propositions below hold under the general $(\beta,\delta)$ structure for any invertible $A_0$.

**Proposition 4** (Flow persistence)**.** *Holding fixed* $(\beta,\delta,A_0,\Lambda)$*, the multiplier matrix* $\mathcal{M}$ *is increasing in* $\phi$ *whenever* $\beta > 0$*.*

This follows directly from (19): the amplification factor $\alpha = \frac{1-\delta\phi}{1-(\beta+\delta)\phi}$ is increasing in $\phi$ whenever $\beta > 0$. The intuition is that more persistent flows are harder to absorb because today's imbalance predicts future imbalances, so the investor demands compensation not just for the current position but for its expected continuation.

**Proposition 5** (Investor horizon)**.** *Holding fixed* $(\beta,\phi,A_0,\Lambda)$ *with* $\phi > 0$ *and* $\beta > 0$*, the multiplier matrix* $\mathcal{M}$ *is increasing in* $\delta$*.*

Again from (19), $\alpha = \frac{1-\delta\phi}{1-(\beta+\delta)\phi}$, and $\partial\alpha/\partial\delta > 0$ whenever $\phi > 0$ and $\beta > 0$. This is a new prediction relative to the myopic benchmark. A more forward-looking investor (larger $\delta$) anticipates the future imbalances that persistent flows generate, so the feedback between flows and expected prices operates over more horizons and raises the multiplier. When $\delta = 0$, only the one-period-ahead expected price matters, and persistent flows amplify the multiplier through $\beta\phi$ alone. When $\delta > 0$, the entire expected path of future prices feeds back into current demand, strengthening the amplification.

Quantitatively, the amplification factor satisfies

$$\alpha(\beta,\delta,\phi) = \frac{1-\delta\phi}{1-(\beta+\delta)\phi} = \frac{1}{1-\dfrac{\beta\phi}{1-\delta\phi}},$$

so adjustment costs effectively raise the discount-rate loading from $\beta\phi$ to $\beta\phi/(1-\delta\phi)$. The "critical persistence" at which the multiplier diverges is $\phi^* = 1/(\beta+\delta)$, which is lower than the myopic threshold $1/\beta$. Forward-looking investors are therefore more sensitive to flow persistence than myopic investors.

**Proposition 6** (Own-price impact and return variance)**.** *Under the* $(\beta,\delta)$ *parameterization, the own-price multiplier for asset* $j$ *is*

$$\mathcal{M}_{jj} = \alpha(\beta,\delta,\phi)\left[(A_0)^{-1}\Lambda\right]_{jj}.$$

*Suppose investors are risk averse, so higher future return variances and covariances weakly reduce demand, and suppose future risk moves inversely with prices, consistent with the so-called*

*leverage effect of Black (1976).*[6] *Then the variance pass-through channel lowers the instrumented elasticity and raises the price multiplier relative to an otherwise identical economy with no risk pass-through. Appendix A states the corresponding matrix sign conditions formally. In the mean–variance benchmark, this reduces to the statement that the own-price impact of a flow shock in asset $j$ is increasing in that asset's return variance.*

See Appendix A for the proof. The first statement follows from the closed-form multiplier in (19). The second is a comparative-static implication of the accounting decomposition in (15). Under mean–variance demand (Appendix B), it specializes to $\mathcal{M}_{jj} = \alpha\,\gamma\beta\,\Sigma_{jj}/\bar{P}_j$.

**Proposition 7** (Systematic versus idiosyncratic directions)**.** *Under the $(\beta,\delta)$ parameterization, let $v$ be a unit eigenvector of $(A_0)^{-1}\Lambda$ with eigenvalue $\lambda$. A unit flow innovation in direction $v$ generates price response*

$$\mathcal{M}\,v = \alpha(\beta,\delta,\phi)\,\lambda\,v.$$

*Hence percentage price impact is larger along directions associated with larger eigenvalues of $(A_0)^{-1}\Lambda$.*

This follows from (19), since $\mathcal{M} = \alpha\,(A_0)^{-1}\Lambda$ and the eigenvectors of $(A_0)^{-1}\Lambda$ diagonalize the multiplier. A flow along a high-eigenvalue direction loads on a combination of assets that exposes the investor to more risk, so it commands a larger price response; an idiosyncratic direction with a low eigenvalue is easier to absorb and moves prices less. Under mean–variance demand, $(A_0)^{-1}\Lambda = \gamma\beta\,\mathrm{Diag}(\bar{P})^{-1}\Sigma$, so the eigenvalues coincide with those of the (price-normalized) covariance matrix and the result reduces to the statement that systematic risk directions have larger multipliers.

[6] As Dew-Becker, Giglio, and Molavi (2025) explain, "leverage effect" is something of a misnomer: the empirical regularity is simply that measured volatility tends to rise when prices fall, regardless of whether financial leverage is the operative mechanism.

**Proposition 8** (Surprise versus expected flows)**.** *Suppose we decompose the one-period change in outside position into surprise and expected components:*

$$\Delta z_{t+1} = \underbrace{\varepsilon_{t+1}}_{\equiv \Delta z^{surp}_{t+1}} + \underbrace{[-(1-\phi) z_t + \eta_t]}_{\equiv \Delta z^{exp}_{t+1}}$$

*Then, if we estimate a regression of log return on them:*

$$\Delta p_{t+1} = a + B_{surp}\, \Delta z^{surp}_{t+1} + B_{exp}\, \Delta z^{exp}_{t+1} + u_{t+1}$$

*The coefficient matrices will be:*

$$\begin{aligned} B_{surp} &= \mathcal{M} \\ B_{exp} &= \mathcal{M} \cdot \underbrace{\left[W + \left(1 - \tfrac{\beta}{1-\delta\phi}\right)\Sigma_\eta\right] \cdot \left[W + \Sigma_\eta\right]^{-1}}_{\textit{attenuation factor}} \end{aligned}$$

*where* $W \equiv \frac{1-\phi}{1+\phi}(\Sigma_\varepsilon + \Sigma_\eta)$.

To gain intuition, consider the special case where $\Sigma_\varepsilon = \sigma^2_\varepsilon I_N$ and $\Sigma_\eta = \sigma^2_\eta I_N$. Then the attenuation factor becomes the scalar

$$\frac{\frac{1-\phi}{1+\phi}(\sigma^2_\varepsilon + \sigma^2_\eta) + \left(1 - \frac{\beta}{1-\delta\phi}\right)\sigma^2_\eta}{\frac{1-\phi}{1+\phi}(\sigma^2_\varepsilon + \sigma^2_\eta) + \sigma^2_\eta},$$

which is less than 1. In this case, the price impact of surprise flow recovers the true multiplier $\mathcal{M}$ while the expected flow commands a smaller multiplier. When $\delta = 0$, the attenuation factor reduces to $(W + (1-\beta)\sigma^2_\eta)/(W + \sigma^2_\eta)$, recovering the myopic benchmark.

**Discussion.** These comparative statics are exactly what make instrumented elasticities context-dependent but still useful. The multiplier matrix is not universal; it depends on risk, persistence, the investor's forward-looking horizon, and the direction and timing of the shock. But conditional on those features, it is the correct object for mapping percentage-of-shares demand shocks into percentage price responses.

Proposition 8 states that surprise flows matter more than anticipated flows for realized price changes because anticipated demand is already capitalized into prices before the flow is realized. Only the unexpected component generates a contemporaneous pricing error that has to be cleared on impact. The coefficient $\frac{\beta}{1-\delta\phi}$ in the attenuation factor captures how strongly forward-looking demand capitalizes expected innovations: the larger $\beta$ and $\delta$ are, the more prices respond to announcements, and the more attenuated the expected-flow coefficient becomes relative to $\mathcal{M}$.

Proposition 5 is a new prediction that distinguishes our framework from the myopic benchmark. Unlike the other four propositions, which are tested directly in Section 5, Proposition 5 does not lend itself to a simple cross-sectional regression because $\delta$ is not directly observable. Indirect evidence comes from the recovery analysis in Section 6: the geometric-decay specifications that allow $\delta > 0$ fit the impulse-response data better than the one-step model that restricts $\delta = 0$, consistent with the presence of forward-looking demand. If the investor population includes agents with adjustment costs or hedging motives ($\delta > 0$), then the amplification of persistent flows is stronger than a myopic model would predict. Conversely, estimating the structural elasticity $A_0$ from the inverse multiplier without accounting for $\delta$ would overstate the true demand sensitivity.

# 5 Empirical Tests of Multiplier Comparative Statics

In this section, we use two demand shocks in the previous literature to test the predictions in Section 4. We find that price multipliers indeed vary as predicted, which suggests that there are empirical regularities in price impacts that are generalizable.

## 5.1 Data

We use two measures of demand that prior papers have argued reflect trading that is unrelated to cash flow information. For more details about their construction and the arguments for why they are unrelated to cash flows, please see the original papers. In our main specification, we compute weekly versions of these variables.

**Flow-induced trading (FIT).** Our first demand measure is flow-induced trading (FIT) introduced in Lou (2012), which captures the non-discretionary stock-level trading of mutual funds and ETFs in response to flows.[7] The measure relies on a specific mechanism: in response to inflows (outflows), funds tend to scale up (down) their pre-existing holdings in a non-discretionary fashion (Coval and Stafford, 2007; Lou, 2012). For example, if Apple's existing weight is 5% in a fund's portfolio, a $1 inflow (outflow) induces the fund to allocate about five more (fewer) cents to Apple. This behavior is true not only for index funds, but also, on average, for active mutual funds and ETFs (see Figure A4 in Li, 2022). Frazzini and Lamont (2008) and Lou (2012) show that the price impact of FIT reverts over time, which suggests that such trades are not motivated by superior cash flow information.

While other researchers have used monthly or quarterly FIT (e.g. Lou, 2012; Ben-David, Li,

[7] While not initially intended as a demand instrument, it has been used as such by subsequent work (e.g. Li, 2022; Chaudhry, 2025; Van der Beck, 2025).

Rossi, and Song, 2022), to obtain higher frequency measures, we use the weekly fund flows from TrimTabs and merge it with 13F fund holdings data from Thomson Reuters. We compute FIT for stock $i$ in week $t$ as:

$$FIT_{i,t} = \frac{\sum_{\text{fund } j} \text{SharesHeld}_{i,j,t-1} \cdot \text{flow}_{j,t}}{\text{SharesOutstanding}_{i,t-1}}$$

where $\text{flow}_{j,t}$ is the net flow in fund $j$ as a fraction of its lagged assets, $\text{SharesHeld}_{i,j,t-1}$ is the lagged number of shares of stock $i$ held by fund $j$, and the denominator is the lagged number of shares outstanding for stock $i$. Normalizing by shares outstanding ensures that when we regress stock returns on FIT, the resulting coefficient is in "price multiplier" units (Gabaix and Koijen, 2022).

**Order flow imbalance (OFI).** Our second demand shock is the Lee-Ready signed order flow imbalance (OFI). This measure takes all trades in the U.S. stock market and classifies them as buyer- or seller-initiated using the Lee and Ready (1991) algorithm. We download daily Lee-Ready signed OFI data from WRDS intraday indicators, aggregate to weekly frequencies, and then normalize by lagged shares outstanding.[8] The OFI measure has been used heavily in the microstructure literature to study price impact at daily or higher frequencies. Recently, Li and Lin (2023) argue that OFI can also be useful for studying price impact at "asset pricing frequencies" (i.e. monthly or lower).

The main benefit of OFI is that it provides much more variation than FIT. While FIT only captures a subset of trades by mutual funds and ETFs, OFI captures all aggressively executed trading in the U.S. stock market and thus has much more variation. The main drawback is the relative lack of knowledge about the drivers of OFI. Li and Lin (2023) conduct extensive tests and

[8] OFI only captures trades that are executed *aggressively*. Many sophisticated institutional investors tend to execute trades slowly and passively to reduce price impact; OFI does not capture those trades.

| | | | | Percentiles | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | Mean | StDev | 1% | 5% | 25% | 50% | 75% | 95% | 99% |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Return | 4,486 | 0.25 | 8.54 | -22.24 | -11.71 | -3.30 | -0.00 | 3.23 | 12.82 | 28.57 |
| FIT | 3,555 | 0.01 | 0.08 | -0.23 | -0.13 | -0.03 | 0.00 | 0.04 | 0.15 | 0.27 |
| OFI | 4,431 | -0.05 | 0.68 | -2.63 | -0.93 | -0.20 | -0.02 | 0.15 | 0.80 | 2.03 |
| Idiosyncratic vol (63 day) | 4,492 | 3.22 | 2.20 | 0.71 | 0.98 | 1.69 | 2.61 | 4.08 | 7.57 | 11.23 |
| Idiosyncratic vol (252 day) | 4,492 | 3.28 | 2.03 | 0.79 | 1.09 | 1.81 | 2.74 | 4.20 | 7.33 | 10.15 |

**Table 2. Summary Statistics.**
We report summary statistics of the main weekly variables in percent. The demand measures are expressed as fraction of shares outstanding. FIT is flow-induced trading as defined in Lou (2012) and computed using weekly flow data from TrimTabs, and OFI is the order flow imbalance measure used in Li and Lin (2023). The sample for return and OFI spans 1993 to 2022, while the sample for FIT starts from 2007 due to availability of weekly fund flow data. Column (1) reports the average number of stocks per period. Idiosyncratic return is defined as stock returns purged of market and industry components, and the reported volatility is for daily returns.

do not find evidence that OFI is related to various measures of cash-flow news. However, they also admit that, due to the difficulty of measuring news, one cannot be fully certain that OFI is unrelated to cash flow news.

Table 2 summarizes our variables. FIT, which is available starting from 2007 due to availability of TrimTabs flow data, has a mean around zero and a standard deviation of 0.08%. This means that the typical FIT realization for a stock in a week amounts to buying or selling 0.08% of shares outstanding of a stock. OFI, which is available from 1993, has more variation: the weekly standard deviation is 0.71% which is almost an order of magnitude larger than that of FIT.

## 5.2 Own-stock multiplier increases with return variance

We test the prediction that price impact increases with return variance (Proposition 6). For each stock $i$ and week $t$, we use the previous 21 (or 63) trading days, which corresponds to roughly a month (or a quarter), to compute the rolling idiosyncratic return volatility, where idiosyncratic returns are obtained via time-series regressions of stock returns on market and Fama-French industry returns.

We then estimate panel regressions of stock returns on flows interacted with lagged volatilities:

$$r_{i,t} = b_0 \cdot f_{i,t} + b_1 \cdot f_{i,t} \times \sigma_{i,t-1} + \tau_t + \tau_i + \epsilon_{i,t}$$

where the controls include time and stock fixed effects. We cluster standard errors by time and stock. We aim to test whether $b_1 > 0$, that is, whether multipliers are larger when return volatilities are higher.

The regression results are reported in Table 3. Columns (1)-(2) show results for FIT where they differ by the lookback window for computing volatilities. The results indicate that multipliers indeed increase with return volatility. For each 1 standard deviation increase in volatility (0.34), price multiplier increases by around $0.34 \times 6.92 \approx 2.35$ to $0.34 \times 8.75 \approx 2.98$, which is sizeable when compared to the multiplier for an average stock (e.g., per column (1), $1.32 + 0.48 \times 6.92 \approx 4.64$). The results are also statistically significant. Corresponding OFI results in columns (5)-(6) are qualitatively consistent.

To examine whether the effect is monotonic in volatilities, we also sort stocks by $\sigma_{i,t-1}$ in each cross-section into quintiles and interact $f_{i,t}$ with those quintile indicators. The results are reported in columns (3)-(4) and (7)-(8). The conclusion is similar and demonstrates that the effect is monotonic. Overall, the results are consistent with the prediction in Proposition 6.

## 5.3 Systematic flows have larger multipliers than idiosyncratic flows

To test the prediction in Proposition 7 that systematic demand variation has larger price impact, we decompose flows into systematic and idiosyncratic components. Concretely, we use cross-sectional regressions to project stock-level flows ($f_{i,t}$) onto 14 commonly used stock characteristics

| Dependent variable: stock return $r_{i,t}$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $f_{i,t}$ = | FIT | | | | OFI | | | |
| Vol window | 1m | 1q | 1m | 1q | 1m | 1q | 1m | 1q |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $f_{i,t}$ | 1.32** | 0.60 | 1.83*** | 1.84*** | 1.07*** | 0.68*** | 1.22*** | 1.18*** |
| | (0.66) | (0.84) | (0.44) | (0.47) | (0.04) | (0.05) | (0.03) | (0.03) |
| $f_{i,t} \times \sigma_{i,t-1}$ | 6.92*** | 8.75*** | | | 2.43*** | 3.02*** | | |
| | (1.68) | (2.17) | | | (0.09) | (0.10) | | |
| $f_{i,t} \times I_{\sigma_{i,t-1}\text{ bin=2}}$ | | | 1.04*** | 0.79*** | | | 0.50*** | 0.41*** |
| | | | (0.24) | (0.25) | | | (0.02) | (0.02) |
| $f_{i,t} \times I_{\sigma_{i,t-1}\text{ bin=3}}$ | | | 1.86*** | 1.87*** | | | 1.03*** | 0.87*** |
| | | | (0.37) | (0.42) | | | (0.03) | (0.03) |
| $f_{i,t} \times I_{\sigma_{i,t-1}\text{ bin=4}}$ | | | 3.26*** | 3.20*** | | | 1.65*** | 1.56*** |
| | | | (0.56) | (0.57) | | | (0.04) | (0.04) |
| $f_{i,t} \times I_{\sigma_{i,t-1}\text{ bin=5}}$ | | | 5.19*** | 5.91*** | | | 2.52*** | 2.80*** |
| | | | (0.93) | (1.11) | | | (0.06) | (0.06) |
| Time FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Stock FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Obs | 2,249,682 | 2,236,975 | 2,249,682 | 2,236,975 | 7,424,584 | 7,333,328 | 7,424,584 | 7,333,328 |
| $R^2$ | 0.177 | 0.178 | 0.177 | 0.178 | 0.156 | 0.158 | 0.153 | 0.155 |
| Within $R^2$ | 0.002 | 0.002 | 0.002 | 0.002 | 0.058 | 0.060 | 0.055 | 0.057 |

**Table 3. Price Impact by Return Volatility.**
We estimate panel regressions of weekly stock return $r_{i,t}$ on contemporaneous flow $f_{i,t}$ and its interactions with lagged annualized idiosyncratic volatility. Columns (1)-(4) and (5)-(8) report results for FIT and OFI, respectively. Columns (1)-(2) and (5)-(6) estimate effects of flows interacted with the value of lagged volatility, while the other columns estimate effects of flows interacted with quintile indicators of lagged return volatility sorted in each cross-section. Odd (even) columns estimate return volatility using 1 month (1 quarter) lookback windows. All regressions control for time and stock fixed effects. Standard errors are clustered by time and stock, and they are reported in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

downloaded from Chen and Zimmermann (2021) and Fama-French 12 industry dummy indicators.[9]

[9]The characteristics include accruals, asset growth, beta, book-to-market ratio, gross profitability, industry momentum, intermediate momentum, 1-year issuance, 5-year issuance, momentum, seasonal momentum, net operating assets, realized volatility, and size. Following common practice (e.g. Kelly, Pruitt, and Su, 2019), we transform each characteristic into a uniform distribution from −0.5 to 0.5 by rank in each cross-section.

For each time period $t$, we estimate realized demand factors $\{g_{k,t}\}_{k=1}^{K}$ via cross-sectional regressions

$$f_{i,t} = \underbrace{a_t + \sum_{k=1}^{K} b_{i,k,t-1} \cdot \hat{g}_{k,t}}_{\equiv f_{i,t}^{\text{systematic}}} + \underbrace{\hat{u}_{i,t}}_{\equiv f_{i,t}^{\text{idiosyncratic}}},$$

which we then use to decompose flows into systematic and idiosyncratic components. We then estimate price multipliers using panel regressions with time fixed effects:

$$r_{i,t} = M_{\text{systematic}} \cdot f_{i,t}^{\text{systematic}} + M_{\text{idiosyncratic}} \cdot f_{i,t}^{\text{idiosyncratic}} + \tau_t + v_{i,t}$$

and cluster standard errors by time and stock. Proposition 7 implies that $M_{\text{systematic}} > M_{\text{idiosyncratic}}$.

The regression results are reported in Table 4. Columns (1) and (2) report results based on the full sample for FIT and OFI, respectively. For both of them, we find that the systematic component-multiplier is significantly higher than the idiosyncratic component-multiplier, as indicated by the last row in the table. In columns (3) to (5), we further split OFI, which has more statistical power than FIT, into three subperiods and estimate results separately. We find that the conclusion is robust across subperiods. Overall, the evidence is consistent with the prediction that systematic demand leads to larger price impact.[10]

## 5.4 More surprising flows have larger multipliers

We now test Proposition 8 which states that surprise flows have larger price impact than expected flows. In theory, a flow is either expected or unexpected. In practice, if investors have heterogeneous information sets, then there can also be intensive margin variation: some flows may be expected

[10] Li and Lin (2023) arrive at a similar conclusion using different specifications.

| Dependent variable: stock return $r_{i,t}$ | | | | | |
|---|---|---|---|---|---|
| $f_{i,t}$ = | FIT | OFI | | | |
| Sample | Full | Full | 1993-2002 | 2003-2012 | 2013 - 2022 |
| | (1) | (2) | (3) | (4) | (5) |
| $f_{i,t}^{\text{systematic}}$ | 6.39*** | 4.40*** | 5.29*** | 3.36*** | 3.94*** |
| | (1.18) | (0.25) | (0.45) | (0.25) | (0.38) |
| $f_{i,t}^{\text{idiosyncratic}}$ | 3.49*** | 2.55*** | 2.92*** | 2.22*** | 1.94*** |
| | (0.35) | (0.04) | (0.06) | (0.06) | (0.05) |
| Time fixed effect | Y | Y | Y | Y | Y |
| Obs | 2,253,935 | 7,475,646 | 3,156,053 | 2,402,155 | 1,917,438 |
| $R^2$ | 0.176 | 0.147 | 0.122 | 0.183 | 0.170 |
| $R^2$ of $f_{i,t}$ | 0.002 | 0.050 | 0.063 | 0.039 | 0.031 |
| $f_{i,t}^{\text{systematic}} - f_{i,t}^{\text{idiosyncratic}}$ | 2.90*** | 1.85*** | 2.37*** | 1.14*** | 2.00*** |
| | (1.07) | (0.24) | (0.44) | (0.25) | (0.38) |

**Table 4. Price Impact for Systematic versus Idiosyncratic Flows.**
For each demand measure, we decompose demand into an aggregate component ($f_{i,t}^{\text{systematic}}$) that corresponds to industry- and stock characteristics-related demand, and a residual ($f_{i,t}^{\text{idiosyncratic}}$). We then use panel regressions to estimate the price multiplier associated with these demand components, controlling for time fixed effect and clustering standard errors by time and stock. Column (1) reports results for FIT. Columns (2)-(5) report results for OFI, with column (2) using the full sample and the other columns using subsamples. The last row of the table reports the multiplier difference with standard errors computed using the Delta method. Standard errors are reported in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

to a higher or lower fraction of investors. Therefore, we decompose realized flow $f_{i,t}$ into three components with decreasing degree of "surprisingness":

$$f_{i,t} = f_{i,t}^{\text{surp}} + f_{i,t}^{\text{less surp}} + f_{i,t}^{\text{predicted}}.$$

To construct these components, we fit two statistical models that predict flows using past data, and the two models differ in the amount of input variables. The more granular model, whose prediction we denote $\hat{f}_{i,t}^{\text{granular}}$, uses up to 13 weekly lags of past demand and returns (the period of approximately a quarter). The coarse model ($\hat{f}_{i,t}^{\text{coarse}}$) is based on less data: 4 weekly lags of past

demand and returns (approximately a month). Then, we define the components as:

$$\begin{aligned} f_{i,t}^{\text{surp}} &= f_{i,t} - \hat{f}_{i,t}^{\text{granular}} \\ f_{i,t}^{\text{less surp}} &= \hat{f}_{i,t}^{\text{granular}} - \hat{f}_{i,t}^{\text{coarse}} \\ f_{i,t}^{\text{predicted}} &= \hat{f}_{i,t}^{\text{coarse}} \end{aligned} \tag{22}$$

To achieve better predictive power than linear models, we use the XGBoost (Extreme Gradient Boosting) model, which takes into account nonlinear interactions between input variables (Chen and Guestrin, 2016). In short, the model builds an ensemble of prediction trees, with each incremental decision tree trained on the residual prediction errors of the existing trees. This is an off-the-shelf supervised learning model commonly used in the machine learning community and we refer the reader to standard online documents for further details.[11]

To approximate the belief of investors who learn in real time, in each year $y$, we use data in years $y-4,\ldots,y-1$ to build a model and compute predictions on year $y$.[12] In unreported tests, we verify that the out-of-sample forecasts are well-calibrated, and that the model performance increases as more lagged variables are added as input variables.

Having decomposed flows as in (22), to test whether more surprising demand has larger price impact, we estimate multipliers via regression

$$r_{i,t} = b_{\text{surp}} \cdot f_{i,t}^{\text{surp}} + b_{\text{less surp}} \cdot f_{i,t}^{\text{less surp}} + b_{\text{predicted}} \cdot f_{i,t}^{\text{predicted}} + \tau_t + \text{Controls}_{i,t} + \epsilon_{i,t},$$

where $\tau_t$ denotes period-specific effects and we control for 13 weekly lags of past returns to control

[11] For instance, see https://xgboost.readthedocs.io/en/stable/.

[12] For learning parameters, we use a max tree depth of 3, a learning rate of 0.1, and 100 rounds of boosting.

for short-term reversal and momentum patterns. We can estimate this either using panel regression and cluster standard errors by time and stock, or using Fama-MacBeth regressions.

The results are reported in Table 5. Columns (1)-(2) report results based on FIT. As predicted by theory, $f_{i,t}^{\text{surp}}$ is associated with larger price multipliers. Panel B reports the pairwise multiplier differences. FIT-based inference has less data and smaller statistical power. While we find statistically significant differences between the coefficients of $f_{i,t}^{\text{surp}}$ and others, we do not discern a significant difference between $f_{i,t}^{\text{less surp}}$ and $f_{i,t}^{\text{predicted}}$ components.

Columns (3)-(4) report results based on OFI. As discussed earlier, OFI has more variation and higher explanatory power for returns, as shown in the last row of Panel A: OFI demand can explain around 5% of weekly return variation as opposed to less than 1% in the case of FIT. Based on the panel regression in column (3), surprise demand $f_{i,t}^{\text{surp}}$ is associated with a price multiplier of 3.07, while the less surprising component $f_{i,t}^{\text{less surp}}$ has a multiplier of 0.45, and the predicted component $f_{i,t}^{\text{predicted}}$ has a multiplier that is indistinguishable from zero. The pairwise multiplier comparisons in Panel B are all statistically significant at the 1% level. The Fama-MacBeth results in column (4) are qualitatively similar. Overall, the results are consistent with the idea that more surprising demand has larger per-unit price impact.

## 5.5 Multiplier increases with flow persistence

Proposition 4 in Section 4 states that the price multiplier scales up with the persistence of flows. To test this, we measure the persistence of FIT and OFI positions, once the expected component is extracted. We then estimate a regression to see if price multipliers increase with persistence.

| Panel A: price impact regressions | | | | |
|---|---|---|---|---|
| | Dependent variable: stock return $r_{i,t}$ | | | |
| $f_{i,t}$ = | FIT | | OFI | |
| | Panel | Fama-MacBeth | Panel | Fama-MacBeth |
| | (1) | (2) | (3) | (4) |
| $f_{i,t}^{\text{surp}}$ | $4.09^{***}$ | $3.95^{***}$ | $3.07^{***}$ | $2.76^{***}$ |
| | (0.50) | (0.24) | (0.06) | (0.03) |
| $f_{i,t}^{\text{less surp}}$ | $1.98^{**}$ | $1.45^{**}$ | $0.45^{***}$ | $0.68^{***}$ |
| | (0.87) | (0.66) | (0.13) | (0.11) |
| $f_{i,t}^{\text{predicted}}$ | $2.12^{**}$ | $1.69^{*}$ | $-0.07$ | $0.11^{*}$ |
| | (1.07) | (0.86) | (0.10) | (0.06) |
| Controls | Y | Y | Y | Y |
| Obs | 1,415,002 | 1,415,002 | 5,573,138 | 5,573,138 |
| $R^2$ | 0.137 | 0.143 | 0.181 | 0.184 |
| Marginal $R^2$ of $f_{i,t}$ | 0.002 | 0.006 | 0.048 | 0.050 |
| Panel B: Coefficient differences | | | | |
| | (1) | (2) | (3) | (4) |
| $f_{i,t}^{\text{surp}} - f_{i,t}^{\text{less surp}}$ | $2.11^{**}$ | $2.50^{***}$ | $2.61^{***}$ | $2.09^{***}$ |
| | (0.92) | (0.65) | (0.15) | (0.12) |
| $f_{i,t}^{\text{less surp}} - f_{i,t}^{\text{predicted}}$ | $-0.13$ | $-0.25$ | $0.52^{***}$ | $0.56^{***}$ |
| | (1.06) | (0.86) | (0.11) | (0.09) |
| $f_{i,t}^{\text{surp}} - f_{i,t}^{\text{predicted}}$ | $1.98^{*}$ | $2.26^{***}$ | $3.13^{***}$ | $2.65^{***}$ |
| | (1.12) | (0.86) | (0.12) | (0.08) |

**Table 5. Price Impact for Surprise versus Expected Flows.**
We use XGBoost machine learning algorithm to decompose flow $f_{i,t}$ into components that are surprising, less surprising, and predicted. We then estimate their price impacts by regressing stock returns on those components, controlling for 13 lags of past weekly returns. Columns (1)-(2) present results for FIT and columns (3)-(4) present results for OFI. Columns (1) and (3) estimate panel regressions with time fixed effects, with standard errors clustered by time and stock. Columns (2) and (4) estimate Fama-MacBeth regressions. Panel A reports the regression coefficients and its last row reports the marginal $R^2$ attributed to demand variables. Panel B reports pairwise coefficient comparisons where, in the case of panel regressions, the standard errors are computed using the Delta method. Standard errors are reported in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

**Constructing $Z_{i,t}^{\text{surp}}$ and $\rho_i$.** Let $f_{i,t}$ be the OFI or FIT flow at time $t$. We estimate the position level by summing up all flows from the beginning of the data sample:

$$Z_{i,t} = \sum_{\tau \le t} f_{i,\tau}$$

To remove the expected component, we first detrend the cumulative position series ($Z_{i,t}$) by stock to isolate the surprise component. Specifically, starting from $Z_{i,t}$ we obtain $Z^{\text{det}}_{i,t}$ by either a cubic LSQ spline (*Spline*) or a flexible Hodrick–Prescott filter (*HP*). We then model the detrended positions for each stock $i$ as a stationary AR(1) in levels with a residual $Z^{\text{surp}}_{i,t}$:

$$Z^{\text{det}}_{i,t} = \rho_i \, Z^{\text{det}}_{i,t-1} + Z^{\text{surp}}_{i,t}.$$

This is estimated by OLS without an intercept on all valid $(t-1,t)$ pairs for stock $i$. Given $\widehat{\rho}_i$, the *position surprise* $Z^{\text{surp}}_{i,t}$ isolates the unanticipated component of the detrended position at $t$. Intuitively, $\widehat{\rho}_i \, Z^{\text{det}}_{i,t-1}$ is the expected continuation of the position based on its own persistence, while $Z^{\text{surp}}_{i,t}$ is the innovation that theory predicts should move prices on impact. These objects ($Z^{\text{surp}}_{i,t}$ and $\widehat{\rho}_i$) enter the regression in interaction to test whether price impact scales with persistence.

**Testing whether price multipliers are an increasing function of persistence.** To bring this prediction to the data in a transparent and model-light way, we use a reduced-form parameterization in which the (conditional) multiplier is allowed to depend linearly on $\rho_i$, written as $(\beta_0 + \beta_1 \rho_i)$, where $\beta_0$ and $\beta_1$ are freely estimated parameters. We estimate the regression:

$$r_{i,t} \;=\; \alpha_t \;+\; \beta_0 \, Z^{\text{surp}}_{i,t} \;+\; \beta_1 \left( Z^{\text{surp}}_{i,t} \times \rho_i \right) \;+\; \varepsilon_{i,t}, \tag{23}$$

where $\alpha_t$ are time fixed effects and standard errors are clustered by time. The interaction term $Z^{\text{surp}}_{i,t} \times \rho_i$ tests whether the flow-induced price impact scales with persistence. We estimate this regression for both weekly FIT and OFI data, using both spline and HP detrending to ensure that the results are not sensitive to the method used to remove low-frequency trends from positions.

| | FIT (Spline) | FIT (HP) | OFI (Spline) | OFI (HP) |
|---|---|---|---|---|
| $Z^{\text{surp}} \times \rho_i$ | $5.782^{*}$ | $27.518^{***}$ | $1.575^{***}$ | $2.324^{***}$ |
| | (3.383) | (9.920) | (0.122) | (0.296) |
| $Z^{\text{surp}}$ | $4.221^{**}$ | $16.166^{***}$ | $1.979^{***}$ | $3.774^{***}$ |
| | (1.722) | (3.982) | (0.072) | (0.096) |
| # Observations | 2,245,961 | 2,245,961 | 7,455,959 | 7,455,959 |
| $R^2$ | 0.004 | 0.001 | 0.039 | 0.012 |

**Table 6. Price Impact and Flow Persistence.**
We estimate the weekly panel regression

$$r_{i,t} = \alpha_t + \beta_0 Z^{\text{surp}}_{i,t} + \beta_1 (Z^{\text{surp}}_{i,t} \times \rho_i) + \varepsilon_{i,t},$$

where $r_{i,t}$ is the return, $Z^{\text{surp}}_{i,t}$ is the position surprise, and $\rho_i$ is the estimated persistence of detrended positions for stock $i$. All specifications absorb time fixed effects and cluster standard errors by time. The interaction term tests whether the price multiplier is increasing in flow persistence, as predicted by Proposition 4. Both FIT (flow-induced trading) and OFI (order flow imbalance) measures are analyzed using cubic spline and Hodrick–Prescott (HP) detrending. Dependent variable is the weekly stock return. Standard errors are reported in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

**Results.** Table 6 reports the results of estimating equation (23) for both weekly FIT and weekly OFI flows using two detrending methods (Spline and HP). Across all specifications, the interaction term $Z^{\text{surp}}_{i,t} \times \rho_i$ is positive and statistically significant, consistent with the model's prediction that more persistent flows generate larger price impacts.

For FIT flows, the coefficients on $Z^{\text{surp}}_{i,t} \times \rho_i$ are large in magnitude (5.8 under spline detrending and 27.5 under the HP filter) and significant at the 10% and 1% levels, respectively. OFI flows exhibit the same qualitative pattern with smaller magnitudes, but the interaction coefficients are strongly statistically significant. Overall, the estimates indicate that price multipliers increase with flow persistence: a one-standard-deviation increase in $\rho_i$ roughly doubles the effect of a flow surprise on returns.

# 6 Recovering Demand in a Dynamic Model

Given that the inverse of an equilibrium multiplier is generally not the contemporaneous structural elasticity, a separate recovery step is needed. In this section, we show that reduced-form price responses do contain enough information to recover the underlying intertemporal demand system. The key observation is that, in the unified log-linear notation, the equilibrium restriction links the structural demand matrices to the reduced-form impulse responses through a linear convolution equation. We state a general recovery result and then develop low-dimensional restrictions that can be taken to the data.

## 6.1 Impulse responses

We begin from the unrestricted equilibrium condition implied by (1) and linearized market clearing:

$$A_0 p_t - \sum_{s=1}^{S} A_s \mu^p_{t,t+s} = \Lambda z_t + \sum_{s=1}^{S} B_s \mu^d_{t,t+s} - \sum_{s=1}^{S} D_s v^r_{t,t+s} + \xi_t. \tag{24}$$

Suppose an identified demand innovation $\varepsilon_t$ induces the impulse-response representations

$$\begin{aligned} z_t &= \bar{z} + \sum_{h=0}^{\infty} \Psi^z_h \varepsilon_{t-h}, \\ p_t &= \bar{p} + \sum_{h=0}^{\infty} \Upsilon_h \varepsilon_{t-h}, \\ \mu^d_{t,t+s} &= \bar{\mu}^d_s + \sum_{h=0}^{\infty} \Phi^d_{s,h} \varepsilon_{t-h}, \qquad s = 1, \ldots, S, \\ v^r_{t,t+s} &= \bar{v}_s + \sum_{h=0}^{\infty} \Omega_{s,h} \varepsilon_{t-h}, \qquad s = 1, \ldots, S. \end{aligned} \tag{25}$$

Here $\Psi_h^z$ is the impulse response of outside holdings, $\Upsilon_h$ is the impulse response of prices, $\Phi_{s,h}^d$ is the impulse response of expected dividends at horizon $s$, and $\Omega_{s,h}$ is the impulse response of the risk state at horizon $s$. These are population infinite-horizon impulse responses; the recovery results below use only the finite subset of horizons that is observed or estimated in practice. Taking expectations in (25) gives

$$\mu_{t,t+s}^p = \mathbb{E}_t[p_{t+s}] = \bar{p} + \sum_{h=0}^{\infty} \Upsilon_{h+s} \varepsilon_{t-h}.$$

Under Assumption 1, $\xi_t$ contributes only its constant component $\bar{\xi}$ to these projections, so it drops out of the coefficient restrictions. Substituting these impulse responses into (24) and matching coefficients on $\varepsilon_{t-h}$ yields

$$A_0 \Upsilon_h - \sum_{s=1}^{S} A_s \Upsilon_{h+s} = \Lambda \Psi_h^z + \sum_{s=1}^{S} B_s \Phi_{s,h}^d - \sum_{s=1}^{S} D_s \Omega_{s,h}, \qquad h \geq 0. \tag{26}$$

Equation (26) is the unrestricted recovery equation. It says that the structural demand parameters must jointly reconcile the observed impulse responses of prices and outside holdings: for each horizon $h$, the price response $\Upsilon_h$ and the outside-holdings response $\Psi_h^z$ are both treated as knowns, and the structural coefficients—current-price elasticity, expected-price elasticity, expected-dividend elasticity, and risk elasticity—are the unknowns. For each $h$, there are $N^2$ equations since this is a $N \times N$ matrix equation. Since this is true for each $h$, we can use finite $H + 1$ (i.e. $h = 0, ..., H$) of these equations to estimate the parameters, leading to a total of $(H + 1)N^2$ equations to estimate the parameters. This is really the basic idea of recovery: with the impulse responses and enough full-rank equations, one can recover the structural parameters with GMM.

## 6.2 Recovery from unrestricted impulse responses

For a given horizon $h$, equation (26) is linear in the unknown coefficient matrices. Stacking the vectorized horizon-by-horizon restrictions yields a system of the form

$$m_H = M_H\, g(\theta), \tag{27}$$

where $m_H$ stacks the observed reduced-form impulse responses through horizon $H$, $g(\theta)$ stacks the structural coefficient matrices implied by the primitive parameter vector $\theta \in \Theta \subset \mathbb{R}^{K_\theta}$, and $M_H$ collects the known coefficient blocks from the vectorized horizon-by-horizon equations. Appendix E gives one explicit construction of this stacked system and the associated counting formulas.

Equation (27) is best viewed as a bookkeeping device. The unrestricted benchmark takes $\theta$ to collect every free element of $(A_0,\ldots,A_S,B_1,\ldots,B_S,D_1,\ldots,D_S)$, so that $g$ is the identity stacking map. Empirical work may instead uses a lower-dimensional parameterization: $\theta$ contains the economically meaningful primitive parameters, while $g(\theta)$ maps those primitives into the full coefficient system.

To write the system in standard GMM notation, define the stacked moment condition

$$\varphi_H(\theta) \equiv m_H - M_H g(\theta).$$

The true parameter $\theta_0$ satisfies $\varphi_H(\theta_0) = 0$.

Equation (26) contributes one $N \times N$ matrix restriction at each horizon, so stacking horizons $h = 0,\ldots,H$ yields $(H+1)N^2$ scalar equations. In the fully unrestricted benchmark, the unknown coefficient vector is very large, especially because each risk block $D_s$ is $N \times m$ with $m = N(N+1)/2$.

So if all $A_s$ parameters are unrestricted, this contributes $(S+1)N^2$ unknowns parameters. If all $B_s$ are unrestricted, this contributes $SN^2$ unknowns. If all $D_s$ are unrestricted, this contributes $SN^2(N+1)/2$ unknowns. In this case, there would be $N^2\left[S(N+5)/2+1\right]$ unknown parameters, and so a necessary condition for identification is that $H \geq S(N+5)/2$. However, more structure can be imposed to lessen the dimensionality of $\theta$ and therefore the burden on the rank condition needed for identification. For instance, if the risk-component of demand depends only on an asset's own conditional variance and $F$ factor variances (which can be formed using linear combinations of $v^r_{t,t+s}$ terms), then each $D_s$ has only $N(1+F)$ free parameters. If the $A_s$ and $B_s$ blocks are unrestricted, then the necessary condition becomes $H \geq 2S + S(1+F)/N$. In short, some additional structure can be imposed to reduce the dimensionality of $\theta$ and the burden on the rank condition needed for identification. Appendix E reports the unrestricted count and some simple special cases.

We can also use cross-sectional aggregation in moment conditions, to obtain less moment conditions than $N^2(H+1)$. While this fully disaggregated system preserves all cross-sectional restrictions, it is high-dimensional and can be potentially noisy in finite samples. Aggregating moments across assets reduces dimensionality, improves numerical stability, and naturally accommodates unbalanced panels by allowing moments to be formed from all available observations at each date. Moreover, by averaging across assets, aggregation mitigates idiosyncratic noise in individual moments, yielding higher signal-to-noise conditions that more efficiently capture the underlying economic restrictions, albeit at the cost of discarding some cross-sectional variation. We provide the recovery result below.

**Proposition 9** (Recovery from parameterized impulse responses)**.** *Suppose Assumption 1 holds, the impulse responses* $\{\Psi^z_h, \Upsilon_h\}_{h=0}^{H+S}$, $\{\Phi^d_{s,h}\}_{s=1,\ldots,S;\, h=0,\ldots,H}$, *and* $\{\Omega_{s,h}\}_{s=1,\ldots,S;\, h=0,\ldots,H}$ *are observed, and let* $\theta_0$ *denote the true parameter vector. If the moment condition* $\varphi_H(\theta) = 0$ *implies* $\theta = \theta_0$*, then* $\theta_0$ *is identified. If* $g$ *is continuously differentiable, a sufficient condition for local identification*

*at* $\theta_0$ *is*

$$\operatorname{rank}\left(\frac{\partial \varphi_H(\theta_0)}{\partial \theta^\top}\right) = K_\theta.$$

This is the standard identification condition for nonlinear GMM problems, so no proof is given. Proposition 9 is deliberately general. The unrestricted system is the special case in which $\theta$ collects every free element of the coefficient matrices. The empirical implementation below does not attempt fully unrestricted recovery, but imposes additional structure. In particular, we impose the return-based restrictions of (5), along with additional structure.

It is important to note that recovery requires a rich time-series structure of shocks and responses. Mathematically, identification becomes easier as the available horizon length $H$ increases, because each additional horizon contributes another block of dynamic restrictions.

This differs from the identification problem emphasized by Binsbergen et al. (2025). Their point is that persistent or predictable instruments generally do *not* identify a ceteris-paribus price elasticity (i.e. $A_0$ here), because they move the entire future path of resale prices. Our point here is different: once the object of interest is the structural *dynamic* demand system, those same persistent responses are useful because they generate the impulse-response variation needed to recover the forward-looking coefficients. For recovery, persistent shocks are therefore a feature, not a bug. And even recovering the contemporaneous ceteris-paribus elasticity $A_0$ would not by itself recover the full dynamic demand system, because counterfactuals in dynamic settings also depend on the continuation coefficients governing responses to future expected returns, dividends, and risk.

For two simple examples of identifying structural parameters in simple cases, the formulas to do so, and more intuition on identification and what the full rank condition means here, we refer the reader to Appendix F.

## 6.3 Recovery with institutional holdings data

Demand system asset pricing is often focused on using institutional-level holdings data, like 13F filings, to recover the demand system at the investor level. We present a model with a representative investor and an empirical implementation that does not use holdings data beyond the construction of the outside-flow measure $z_t$ (e.g. FIT). However, we briefly trace out how the same recovery method can be used to identify investor-specific structural demand coefficients with investor-level holdings data.

Consider investor $i$ with demand

$$q_{i,t} = -A_{0,i}\, p_t + \sum_{s=1}^{S} A_{s,i}\, \mu^{p}_{t,t+s} + \sum_{s=1}^{S} B_{s,i}\, \mu^{d}_{t,t+s} - \sum_{s=1}^{S} D_{s,i}\, v^{r}_{t,t+s} + \xi_{i,t}. \tag{28}$$

Let the same identified demand shock $\varepsilon_t$ induce the additional impulse-response representation

$$q_{i,t} = \bar{q}_i + \sum_{h=0}^{\infty} \Psi^{q_i}_{h}\, \varepsilon_{t-h}.$$

Assume likewise that the investor-specific latent demand shifter $\xi_{i,t}$ is exogenous to $\varepsilon_t$, so its impulse response is zero. Taking impulse responses of (28) with respect to $\varepsilon_t$ and rearranging yields the investor-level recovery equation:

$$A_{0,i}\, \Upsilon_h - \sum_{s=1}^{S} A_{s,i}\, \Upsilon_{h+s} = \sum_{s=1}^{S} B_{s,i}\, \Phi^{d}_{s,h} - \sum_{s=1}^{S} D_{s,i}\, \Omega_{s,h} - \Psi^{q_i}_{h}, \qquad h \geq 0. \tag{29}$$

This is structurally identical to (26), with $-\Psi^{q_i}_{h}$ in place of $\Lambda\Psi^{z}_{h}$ on the right-hand side. The price impulse responses $\Upsilon_h$, $\Phi^{d}_{s,h}$, and $\Omega_{s,h}$ are aggregate objects estimated from price and dividend data exactly as before; the only additional input is $\Psi^{q_i}_{h}$, the impulse response of investor $i$'s log holdings to the demand shock. The same GMM system applies, and the same rank condition recovers the

investor-specific coefficients $(A_{0,i},\ldots,A_{S,i},B_{1,i},\ldots,B_{S,i},D_{1,i},\ldots,D_{S,i})$. This derivation imposes homogeneous beliefs: the aggregate objects $\mu^d_{t,t+s}$ and $v^r_{t,t+s}$ are assumed common across investors, so the aggregate IRFs $\Phi^d_{s,h}$ and $\Omega_{s,h}$ apply to each investor's demand equation.

Consistency with the aggregate equation follows from market clearing. With $I$ investors, total shares satisfy $\sum_i Q_{i,t} + Z_t = \mathbf{1}_N$. Log-linearizing around $(\bar{Q}_i, \bar{Z})$ gives the multi-investor clearing condition:

$$\sum_i \mathrm{Diag}(\bar{Q}_i)\, q_{i,t} + z_t = 0,$$

which in impulse-response form is $\sum_i \mathrm{Diag}(\bar{Q}_i)\, \Psi^{q_i}_h = -\Psi^z_h$. Premultiplying (29) by $\mathrm{Diag}(\bar{Q}_i)$ and summing over $i$ yields

$$\begin{aligned}
&\Big(\sum_i \mathrm{Diag}(\bar{Q}_i) A_{0,i}\Big)\Upsilon_h - \sum_{s=1}^{S}\Big(\sum_i \mathrm{Diag}(\bar{Q}_i) A_{s,i}\Big)\Upsilon_{h+s} \\
&= \sum_{s=1}^{S}\Big(\sum_i \mathrm{Diag}(\bar{Q}_i) B_{s,i}\Big)\Phi^d_{s,h} - \sum_{s=1}^{S}\Big(\sum_i \mathrm{Diag}(\bar{Q}_i) D_{s,i}\Big)\Omega_{s,h} + \Psi^z_h.
\end{aligned}$$

Premultiplying by $\Lambda = \mathrm{Diag}(\bar{Q})^{-1}$, where $\bar{Q} = \sum_i \bar{Q}_i$, this is exactly (26) with aggregate structural parameters defined as share-weighted averages, e.g. $A_0 \equiv \Lambda \sum_i \mathrm{Diag}(\bar{Q}_i) A_{0,i}$.

## 6.4 Simple parameterizations for empirical implementation

The unrestricted recovery system contains $(H{+}1)N^2$ moment conditions and $N^2[\, S(N{+}5)/2{+}1\,]$ free parameters. In the weekly implementation, we reduce dimensionality along both margins. On the moment side, for each shocked stock $i$ and horizon $h$, we replace the full cross-section with two

moments: the own response and the average-other response,

$$\bar{x}_{-i,t+h} \equiv \frac{1}{N_t - 1} \sum_{j \neq i} x_{j,t+h},$$

so that each shock-horizon pair contributes two equations rather than $N^2$. On the parameter side, we fit the demand system directly in this own/average-other basis, with one coefficient governing the own response and one governing the average-other response at each horizon. These restrictions make the system tractable while preserving the variation most relevant for identification: they accommodate unbalanced panels, pool noisy cross-sectional information, and avoid a high-dimensional GMM problem with potentially many weakly informative moments and parameters.

We then fit parsimonious dynamic demand systems directly in this own/average-other basis. At each horizon, one coefficient governs the own response and one governs the average-other response. This is already more flexible than logit-style demand, where a single parameter governs both own and cross substitution. It is also empirically natural: Chaudhry and Davis (2026) find that investors treat individual stocks as weak substitutes and report quantitatively small cross-price multipliers and cross-price elasticities. We emphasize that this cross-elasticity specification is much more restrictive than the Chaudhry and Davis (2026) and Haddad, He, Huebner, Kondor, and Loualiche (2025a) specifications. In the notation below, we leave $S$ generic; in the implemented weekly recovery reported in Table 8, we set $S = 4$. The reduced-form objects are the impulse responses of future flows, cumulative returns, and weekly total variance in this own/average-other basis:

$$\Psi_h^{\text{diag}}, \quad \Psi_h^{\text{off}}, \qquad \Xi_h^{\text{diag}}, \quad \Xi_h^{\text{off}}, \qquad \Omega_h^{\text{diag}}, \quad \Omega_h^{\text{off}}.$$

**Example 1: naive static recovery.** If demand is purely static and depends only on current prices,

$$q_t = -A_0 p_t + z_t,$$

with no future-return, future-dividend, or risk terms, then the equilibrium condition (24) reduces to $A_0 p_t = \Lambda z_t$. In the paper's empirical normalization—where both FIT and OFI are scaled by lagged shares outstanding so that $\Lambda = I$—this gives

$$p_t = A_0^{-1} z_t.$$

In that static case, the price multiplier is $\mathcal{M} = A_0^{-1}$, so the recovered elasticity is exactly

$$A_0 = \mathcal{M}^{-1}.$$

Equivalently, this is the $S = 0$ (static) special case of the recovery equation: with $\Lambda = I$ and no continuation terms, the recovery equation at $h = 0$ reduces to $A_0 \Xi_0 = I$, i.e. $A_0 = \Xi_0^{-1} = \mathcal{M}^{-1}$. In the reduced own/average-other system, the static benchmark therefore estimates only two parameters: $A_0^{\text{diag}}$ and $A_0^{\text{off}}$. These two terms, after normalization into the aggregate elasticity units used in the empirical section, reduce to the two unique elements of $\mathcal{M}^{-1}$. We denote the corresponding two elements of the price multiplier by $\mathcal{M}^{\text{diag}}$ and $\mathcal{M}^{\text{off}}$ below in Table 8.

**Example 2: one-step recovery in an own/other basis.** Suppose $S = 1$, so that investors respond only to one-period-ahead expected returns. The one-step $C$-system then implies

$$\Psi_h^{\text{diag}} = C_1^{\text{diag}}\left(\Xi_h^{\text{diag}} - \beta\Xi_{h+1}^{\text{diag}}\right) + C_1^{\text{off}}\left(\Xi_h^{\text{off}} - \beta\Xi_{h+1}^{\text{off}}\right),$$
$$\Psi_h^{\text{off}} = C_1^{\text{off}}\left(\Xi_h^{\text{diag}} - \beta\Xi_{h+1}^{\text{diag}}\right) + C_1^{\text{diag}}\left(\Xi_h^{\text{off}} - \beta\Xi_{h+1}^{\text{off}}\right),$$

for $h = 0, \ldots, H$. This restriction has exactly two free return-demand parameters, $(C_1^{\text{diag}}, C_1^{\text{off}})$. If risk is added in the same one-step form, the system becomes

$$\Psi_h^{\text{diag}} = C_1^{\text{diag}}\left(\Xi_h^{\text{diag}} - \beta\Xi_{h+1}^{\text{diag}}\right) + C_1^{\text{off}}\left(\Xi_h^{\text{off}} - \beta\Xi_{h+1}^{\text{off}}\right) - D_1^{\text{diag}}\Omega_{h+1}^{\text{diag}} - D_1^{\text{off}}\Omega_{h+1}^{\text{off}},$$
$$\Psi_h^{\text{off}} = C_1^{\text{off}}\left(\Xi_h^{\text{diag}} - \beta\Xi_{h+1}^{\text{diag}}\right) + C_1^{\text{diag}}\left(\Xi_h^{\text{off}} - \beta\Xi_{h+1}^{\text{off}}\right) - D_1^{\text{off}}\Omega_{h+1}^{\text{diag}} - D_1^{\text{diag}}\Omega_{h+1}^{\text{off}},$$

so the variance-augmented one-step model has four free parameters: $(C_1^{\text{diag}}, C_1^{\text{off}}, D_1^{\text{diag}}, D_1^{\text{off}})$. In either case, each horizon contributes two equations, one diagonal and one average-other, so using $h = 0, \ldots, H$ gives $2(H + 1)$ equations. In the empirical one-step rows of Table 8, we use horizons $h = 0,1,2,3$, so the estimator uses 8 equations.

**Example 3: two-step recovery.** Suppose $S = 2$, so that investors look two periods ahead. A fully unrestricted two-step system would allow separate diagonal/off-diagonal pairs at the two horizons,

$$\text{horizon 1: } (C_1^{\text{diag}}, C_1^{\text{off}}), \qquad \text{horizon 2: } (C_2^{\text{diag}}, C_2^{\text{off}}).$$

For parsimony, and because the data cannot separately pin down four own/other coefficients at two horizons, Table 8 imposes the restricted two-step shape

$$C_2 = \delta C_1,$$

so the estimated two-step model is summarized by three free return-demand parameters:

$$(C_1^{\text{diag}}, C_1^{\text{off}}, \delta).$$

The implied flow equations are

$$\Psi_h^{\text{diag}} = C_1^{\text{diag}}\left(\Xi_h^{\text{diag}} - \beta\Xi_{h+1}^{\text{diag}}\right) + C_1^{\text{off}}\left(\Xi_h^{\text{off}} - \beta\Xi_{h+1}^{\text{off}}\right) + \delta C_1^{\text{diag}}\left(\Xi_{h+1}^{\text{diag}} - \beta\Xi_{h+2}^{\text{diag}}\right) + \delta C_1^{\text{off}}\left(\Xi_{h+1}^{\text{off}} - \beta\Xi_{h+2}^{\text{off}}\right),$$

$$\Psi_h^{\text{off}} = C_1^{\text{off}}\left(\Xi_h^{\text{diag}} - \beta\Xi_{h+1}^{\text{diag}}\right) + C_1^{\text{diag}}\left(\Xi_h^{\text{off}} - \beta\Xi_{h+1}^{\text{off}}\right) + \delta C_1^{\text{off}}\left(\Xi_{h+1}^{\text{diag}} - \beta\Xi_{h+2}^{\text{diag}}\right) + \delta C_1^{\text{diag}}\left(\Xi_{h+1}^{\text{off}} - \beta\Xi_{h+2}^{\text{off}}\right).$$

If risk is added with the same two-step shape, $D_2 = \delta D_1$, then the model has five free parameters:

$$\left(C_1^{\text{diag}}, C_1^{\text{off}}, \delta, D_1^{\text{diag}}, D_1^{\text{off}}\right).$$

The number of equations is again two per horizon, so using $h = 0, \ldots, H$ yields $2(H+1)$ equations in both the return-only and variance-augmented versions. In the empirical two-step rows of Table 8, we use horizons $h = 0,1,2,3,4$, so the estimator uses 10 equations.

**Example 4: geometric continuation in the $C$ coefficients.** Extending the two-step restriction to a finite horizon, the most persistent specification in Table 8 imposes a geometric decay on the

return-sensitivity matrices,

$$C_s = \delta^{s-1} C_1, \qquad s = 1, \ldots, S,$$

and, when risk is included, the same low-dimensional restriction on the risk coefficients,

$$D_s = \delta^{s-1} D_1, \qquad s = 1, \ldots, S.$$

In the same own/other basis, this implies

$$\Psi_h^{\text{diag}} = \sum_{s=1}^{S} \delta^{s-1} \left[ C_1^{\text{diag}} \left( \Xi_{h+s-1}^{\text{diag}} - \beta \Xi_{h+s}^{\text{diag}} \right) + C_1^{\text{off}} \left( \Xi_{h+s-1}^{\text{off}} - \beta \Xi_{h+s}^{\text{off}} \right) - D_1^{\text{diag}} \Omega_{h+s}^{\text{diag}} - D_1^{\text{off}} \Omega_{h+s}^{\text{off}} \right],$$

$$\Psi_h^{\text{off}} = \sum_{s=1}^{S} \delta^{s-1} \left[ C_1^{\text{off}} \left( \Xi_{h+s-1}^{\text{diag}} - \beta \Xi_{h+s}^{\text{diag}} \right) + C_1^{\text{diag}} \left( \Xi_{h+s-1}^{\text{off}} - \beta \Xi_{h+s}^{\text{off}} \right) - D_1^{\text{off}} \Omega_{h+s}^{\text{diag}} - D_1^{\text{diag}} \Omega_{h+s}^{\text{off}} \right].$$

Without risk, the geometric model therefore has three free parameters:

$$(C_1^{\text{diag}}, C_1^{\text{off}}, \delta).$$

With risk, it has five:

$$\left(C_1^{\text{diag}}, C_1^{\text{off}}, \delta, D_1^{\text{diag}}, D_1^{\text{off}}\right).$$

As above, the reduced own/average-other system contributes two equations at each horizon, so using $h = 0, \ldots, H$ gives $2(H+1)$ equations with or without the variance block. In the empirical geometric rows of Table 8, we use horizons $h = 0,1,2,3,4$, so the estimator also uses 10 equations. These four examples correspond exactly to the rows of Table 8 in the empirical results in Section 6.5: naive static $\mathcal{M}^{-1}$, one-step $C$, two-step $C$, and geometric $C$, each with and without the risk augmentation. In the empirical weekly implementation, the geometric row uses the finite truncation $S = 4$. For

the two-step and geometric specifications, the shape parameter $\delta$ is estimated by one-dimensional minimum distance (GMM): conditional on a candidate $\delta$, the diagonal and average-other coefficients are obtained by least squares, and $\delta$ is then chosen to minimize the distance between model-implied and estimated impulse responses. See Appendix G for more details about the recovery estimation. In this appendix, we write down the estimator in detail, including the weighting scheme and the treatment of the discount factor.

**Risk-augmented specifications.** For the risk-augmented specifications, we estimate one additional set of horizon-by-horizon local projections using the same weekly Fama-MacBeth design as (32)–(33):

$$\begin{aligned} \sigma_{i,t+h} &= a_{t,h}^{\sigma,d} + \Omega_h^d \, \varepsilon_{i,t} + \Delta_h^\top X_{i,t} + \gamma_h \, \sigma_{i,t} + u_{i,t+h}^{\sigma,d}, \\ \bar{\sigma}_{-i,t+h} &= a_{t,h}^{\sigma,o} + \Omega_h^o \, \varepsilon_{i,t} + \bar{\Delta}_h^\top X_{i,t} + \bar{\gamma}_h \, \sigma_{i,t} + u_{i,t+h}^{\sigma,o}, \end{aligned} \tag{30}$$

for $h = 0, \ldots, 8$. Here $\sigma_{i,t+h}$ is the stock's weekly total variance at date $t + h$, measured as the sum of squared daily log returns within week $t + h$; $\bar{\sigma}_{-i,t+h}$ is the corresponding cross-sectional average across all stocks other than $i$; $\sigma_{i,t}$ is the same weekly total-variance measure at date $t$ included as an additional level control; and $X_{i,t}$ is the same four-lag control vector as in (32). The coefficients $\Omega_h^d$ and $\Omega_h^o$ are the own and average-other variance impulse responses at horizon $h$; these are the $\hat{\Omega}$ objects used as generated inputs in the second step.

## 6.5 Empirical recovery results

We now implement these restrictions using the same weekly data studied in Section 5. Our main specification uses the overlapping 2007–2018 sample, where both FIT and OFI are observed on the

same stock-weeks. Because both measures are already scaled by lagged shares outstanding, they are directly comparable in the model's units. We therefore combine them into a single weekly flow measure using a variance-weighted convex combination and treat the innovation in that combined flow as the identified weekly demand shock. Appendix G provides more details about the data and estimation procedure for the weekly recovery exercise.

The identified weekly demand shock is the surprise component of the combined flow measure. In the baseline specification we estimate a pooled stock-demeaned AR(4),

$$z_{i,t} = \sum_{\ell=1}^{4} \phi_\ell z_{i,t-\ell} \; + \; \sum_{\ell=1}^{4} \kappa_\ell r_{i,t-\ell} \; + \; \varepsilon_{i,t}, \tag{31}$$

where $z_{i,t}$ denotes the combined weekly flow measure and $r_{i,t}$ is the weekly log return. The residual $\varepsilon_{i,t}$ is our estimated weekly demand shock. We then estimate horizon-by-horizon local projections using weekly Fama-MacBeth cross-sectional regressions:

$$z_{i,t+h} = \alpha^d_{t,h} + \Psi^d_h \varepsilon_{i,t} + \Gamma_h^\top X_{i,t} + u^d_{i,t+h}, \tag{32}$$

$$\bar{z}_{-i,t+h} = \alpha^o_{t,h} + \Psi^o_h \varepsilon_{i,t} + \bar{\Gamma}_h^\top X_{i,t} + u^o_{i,t+h},$$

$$\sum_{j=0}^{h} r_{i,t+j} = a^d_{t,h} + \Xi^d_h \varepsilon_{i,t} + \Theta_h^\top X_{i,t} + v^d_{i,t+h}, \tag{33}$$

$$\overline{\sum_{j=0}^{h} r_{-i,t+j}} = a^o_{t,h} + \Xi^o_h \varepsilon_{i,t} + \bar{\Theta}_h^\top X_{i,t} + v^o_{i,t+h},$$

for $h = 0, \ldots, 8$, where $\bar{z}_{-i,t+h}$ and $\overline{\sum_{j=0}^{h} r_{-i,t+j}}$ denote the cross-sectional average across all stocks other than $i$ in week $t$. The control vector $X_{i,t}$ contains four lags of flows and returns:

$$X_{i,t} \equiv \left(z_{i,t-1}, \ldots, z_{i,t-4}, r_{i,t-1}, \ldots, r_{i,t-4}\right)^\top.$$

| Shock | $\mathcal{M}_{jj}$ | $\mathcal{M}_{j,-j}$ | $\mathcal{M}_{jj}^{-1}$ |
|---|---|---|---|
| Combined weekly shock | 2.268<br>(0.098) | -0.0006<br>(< 0.0001) | 0.441<br>(0.009) |

**Table 7. Weekly Multiplier Matrix in Own/Other Form.**
The table reports the contemporaneous own and average-other price responses to the combined weekly demand shock on the common 2007–2018 overlap sample. The shock is the residual from the stock-demeaned AR(4) in (31), estimated on the variance-weighted combined flow measure. The final column reports the naive inverse-multiplier benchmark used in Table 8. Standard errors are Newey-West across weekly Fama-MacBeth coefficient series.

Including these lags in every horizon-$h$ projection aligns the local projections with the same information set used in the AR(4) forecasting equation, so that $\varepsilon_{i,t}$ is interpreted as the innovation in weekly demand relative to that baseline forecast. The AR(4) baseline is chosen to match the specification used in the rest of the weekly empirical section and to provide a flexible reduced-form forecast at roughly a one-month horizon.

Table 7 reports these reduced-form objects for the combined weekly demand shock on the common 2007–2018 overlap sample. The contemporaneous own-price multiplier is about 2.27, so the corresponding naive inverse benchmark is about 0.44. We call this benchmark "naive" because it shuts down all continuation terms and equates the structural elasticity with the contemporaneous inverse multiplier alone.

Table 8 reports the recovery results for that single combined-shock system on the common 2007–2018 overlap sample. The first row is the naive static benchmark based on $\mathcal{M}^{-1}$ alone. It delivers a diagonal elasticity of only about 0.44, because it effectively assumes away all future expected-return and risk effects. Every dynamic recovery then implies a much larger structural diagonal elasticity.

The one-step $C$-only specification already raises the diagonal coefficient sharply, to about 8.2, and adding risk moments increases it further to about 9.9. This is exactly the direction predicted by the theory: once the shock is allowed to worsen future risk-bearing conditions, more of the observed price response must be attributed to a stronger expected-return elasticity. The same pattern holds

| Spec | $C_1^{\text{diag}}$ | $C_1^{\text{off}}$ | Shape ($\delta$) | $D_1^{\text{diag}}$ | $D_1^{\text{off}}$ |
|---|---|---|---|---|---|
| Naive static $\mathcal{M}^{-1}$ | 0.441<br>(0.009) | 0.0001<br>(< 0.0001) | – | – | – |
| One-step $C$ only | 8.201<br>(1.116) | -0.0003<br>(< 0.0001) | – | – | – |
| One-step $C$ + var. | 9.872<br>(1.844) | -0.0002<br>(0.0001) | – | 12.731<br>(8.560) | 0.0009<br>(0.0004) |
| Two-step $C$ only | 8.164<br>(1.335) | -0.0003<br>(< 0.0001) | < 0.001<br>(0.419) | – | – |
| Two-step $C$ + var. | 8.452<br>(1.853) | -0.0002<br>(< 0.0001) | 0.978<br>(0.357) | 9.905<br>(7.348) | 0.0009<br>(0.0004) |
| Geo. $C$ only | 8.164<br>(1.672) | -0.0003<br>(< 0.0001) | < 0.001<br>(0.340) | – | – |
| Geo. $C$ + var. | 9.295<br>(1.921) | -0.0002<br>(0.0001) | 0.271<br>(0.339) | 10.697<br>(7.141) | 0.0009<br>(0.0004) |

**Table 8. Structural Recovery from Weekly Own/Other Impulse Responses.**
The table reports structural coefficients recovered from the weekly own and average-other impulse responses of the combined demand shock on the common 2007–2018 overlap sample. "Naive static $\mathcal{M}^{-1}$" is the benchmark that ignores all continuation terms beyond the contemporaneous multiplier. The remaining rows estimate dynamic return-based $C$-systems: one-step, two-step, and geometric restrictions, each with and without risk moments based on weekly total-variance impulse responses. The two-step row imposes $S = 2$, while the geometric row uses the finite truncation $S = 4$ with geometric decay. The "Shape" column reports the decay parameter, $\delta$, for the two-step and geometric specifications. Standard errors in parentheses are obtained by resampling weeks.

for the more persistent parameterizations. The two-step specification gives $\hat{C}_1^{\text{diag}} \approx 8.16$ without risk and $\hat{C}_1^{\text{diag}} \approx 8.45$ with it. The geometric specification gives $\hat{C}_1^{\text{diag}} \approx 8.16$ and $\hat{C}_1^{\text{diag}} \approx 9.29$, respectively. Across all six dynamic specifications, the average-other coefficient remains essentially zero.

The weak average-other estimate is consistent with broader evidence that substitution across individual stocks is limited at the horizons relevant for demand-based price impact. Chaudhry and Davis (2026) find similarly weak cross-stock substitution once the object of interest is an asset-level demand shift rather than a broad market-wide reallocation. Our recovery estimates point in the

| Shock | $\Omega_{jj,1}$ | $\Omega_{j,-j,1}$ |
|---|---|---|
| Combined weekly shock | -0.0058 | $< 0.00001$ |
| | (0.0038) | ($< 0.00001$) |

**Table 9. Weekly Volatility Impulse Responses in Own/Other Form.**
The table reports one-week-ahead weekly total-variance responses to the combined weekly demand shock in own and average-other form. These are the variance moments used in the risk augmentation of Table 8. Standard errors are Newey-West across weekly Fama-MacBeth coefficients.

same direction: the dominant structural margin is own-stock demand, and the average offsetting reallocation into other stocks is close to zero. In our specification, separate parameters govern the own and average-other coefficients within each period's reduced system, which is more flexible than logit-style demand, where a single parameter controls both. A single average-other coefficient may appear overly parsimonious, but the near-zero estimates suggest that it is a reasonable approximation for individual stocks. Using a single parameter for the own-price elasticity is standard in the literature; Haddad et al. (2025a), for example, discuss relaxing this restriction.

Table 9 reports the variance moments that discipline the risk augmentation of Table 8, while Table 10 shows the corresponding Black (1976) inverse price-and-variance relationship estimated with the same combined shock. The key fact is that the own one-week-ahead weekly total-variance response is negative, while the contemporaneous return response is positive. Thus the estimated shock raises price on impact but lowers the subsequent variance term, so the implied risk pass-through has the leverage-effect sign. In the recovery system, this extra variance channel absorbs part of the observed flow response and therefore pushes the recovered own-price elasticity upward once it is included explicitly. This covariance/risk effect has been emphasized by He et al. (2025), and empirically we find that it matters substantially.

Two features of these estimates deserve emphasis. First, every dynamic specification delivers an own-price elasticity far above the naive inverse-multiplier benchmark, which is the central prediction of a forward-looking demand system. Second, within each pair of models, adding risk

| Shock | $z_t \to r_t$ | $z_t \to \log(1+\sigma^2_{t+1})$ | $\Pi^v$ |
|---|---|---|---|
| Combined weekly shock | 2.260 | -0.010 | -0.004 |
| | (0.094) | (0.004) | (0.002) |

**Table 10. Black Leverage Effect for the Combined Weekly Shock.**
The table reports reduced-form pass-through estimates for the combined weekly demand shock on the common 2007–2018 overlap sample. The first column gives the contemporaneous return response, the second gives the response of one-week-ahead log total variance controlling for current log total variance, and the third reports the implied variance pass-through per unit of price response, $\Pi^v$. A negative $\Pi^v$ indicates that the shock has the Black leverage-effect sign. Standard errors in parentheses are Newey-West across weekly Fama-MacBeth coefficient series.

moments raises the diagonal coefficient rather than lowering it. The richer dynamic restrictions therefore sharpen the same conclusion rather than overturning it: the weekly data point to a structural own-price elasticity in the neighborhood of eight to ten, with little evidence of average cross-stock substitution in the own/other basis.

## 6.6 Interpretation

Proposition 9 shows why the log-linear rewrite is useful. The same coefficients that describe structural demand in Section 3 and determine equilibrium multipliers in Section 4 are also the objects recovered from reduced-form dynamics. Nothing conceptually new has to be introduced for the recovery section. One only needs enough support in the shock process and enough variation in the price impulse responses to invert the system.

The proposition also highlights a practical limitation. A single contemporaneous multiplier does not generally identify the demand system. Recovery requires the *path* of price responses together with the path of flow responses. This is exactly where our perspective departs from Binsbergen et al. (2025). For estimating a ceteris-paribus elasticity, persistent instruments are problematic because they shift future resale values as well as the contemporaneous price. For recovering the structural dynamic system, however, the relevant object is precisely that propagation mechanism. The more informative and persistent the shock path is, the more dynamic restrictions it generates.

Persistent instruments are therefore not a nuisance here; they are what make the forward-looking demand coefficients recoverable.

# 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 any DSAP design depends on the process generating the price movement. This raises a practical question for counterfactual analysis: if the elasticity is estimated from a shock with one set of attributes (say, a persistent and systematic flow), it may not apply to a counterfactual involving a shock with different persistence, risk profile, or transmission into future prices. For many DSAP applications, the relevant object is therefore the instrumented elasticity associated with the identifying shock, not the structural own-price coefficient that holds all future expected prices, returns, and risk fixed.

We formalize this distinction within a unified log-linear demand system. The instrumented elasticity is a shock-specific weighted combination of structural demand coefficients, and its inverse is the price multiplier for that shock. In a benchmark equilibrium, we derive comparative statics showing that price multipliers are larger when risk is higher, flows are more persistent, shocks load on more systematic directions, and flows are more surprising. The empirical evidence supports all four predictions.

The same framework addresses the recovery problem. Persistent shocks are not a nuisance; they are informative, because the path of price and flow responses contains precisely the variation needed to identify forward-looking demand coefficients. Under standard invertibility conditions, the structural demand system can be recovered from reduced-form impulse responses. In weekly data,

the quantitative gap is large: the naive inverse-multiplier benchmark, which ignores all forward-looking terms and treats the contemporaneous price response as sufficient, implies an own-price elasticity of about 0.4, whereas the recovered structural elasticity in the main combined-shock specification is about 8 to 10. The low naive estimate reflects the persistence of the identifying shock, not intrinsically weak demand: once the path of future expected returns and risk is accounted for, investors are far more responsive to price changes than the reduced-form multiplier alone would suggest. With the right dynamic structure, these structural objects can be recovered and applied to counterfactuals whose shock attributes differ from those of the original identifying variation.

# References


Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2022, Ratings-driven demand and systematic price fluctuations, *The Review of Financial Studies* 35, 2790–2838.

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

Black, Fischer, 1976, Studies of stock price volatility changes, in *Proceedings of the 1976 Meetings of the Business and Economic Statistics Section*, 177–181 (American Statistical Association).

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

Chaudhry, Aditya, 2025, The impact of prices on analyst cash flow expectations: Reconciling subjective beliefs data with rational discount rate variation, *Journal of Financial Economics* 171, 104095.

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

Chen, Andrew Y, and Tom Zimmermann, 2021, Open source cross-sectional asset pricing, *Critical Finance Review, Forthcoming* .

Chen, Tianqi, and Carlos Guestrin, 2016, Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Coval, Joshua, and Erik Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479–512.

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.

Dew-Becker, Ian, Stefano Giglio, and Pooya Molavi, 2025, Learning and the emergence of nonlinearity in financial markets, NBER Working Paper 34584.

Frazzini, Andrea, and Owen A Lamont, 2008, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

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

Gabaix, Xavier, and Ralph SJ Koijen, 2022, In search of the origins of financial fluctuations: The inelastic markets hypothesis, Working paper.

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

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.

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

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.

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

Li, Jiacui, 2022, What drives the size and value factors?, *Review of Asset Pricing Studies* 12, 845–885.

Li, Jiacui, and Zihan Lin, 2023, Prices are less elastic at more aggregate levels, Working paper.

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

Lou, Dong, 2012, A flow-based explanation for return predictability, *The Review of Financial Studies* 25, 3457–3489.

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

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

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

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, Working paper.

# Appendix

## A Proofs

*Proof of Proposition 1.* By linearized market clearing (7),

$$q_t = -\Lambda z_t.$$

Differentiating with respect to $z_t$ gives

$$\frac{dq_t}{dz_t^\top} = -\Lambda.$$

By definition, the price multiplier matrix is

$$\mathcal{M}_t(z_t) = \frac{dp_t}{dz_t^\top}.$$

Invertibility of $\mathcal{M}_t(z_t)$ then allows us to substitute into (12):

$$\mathcal{E}_t(z_t) = -\frac{dq_t}{dz_t^\top}\mathcal{M}_t(z_t)^{-1} = \Lambda\mathcal{M}_t(z_t)^{-1}.$$

This proves (13).

Now define the aggregate demand elasticity matrix

$$\tilde{\mathcal{E}}_t(z_t) \equiv \Lambda^{-1}\mathcal{E}_t(z_t).$$

Using (13),

$$\tilde{\mathcal{E}}_t(z_t) = \Lambda^{-1}\Lambda\mathcal{M}_t(z_t)^{-1} = \mathcal{M}_t(z_t)^{-1},$$

which proves (14). □

*Proof of Proposition 2.* Total differentiation of (1) with respect to $z_t$ gives

$$\frac{dq_t}{dz_t^\top} = A_0\left(-\frac{dp_t}{dz_t^\top}\right) + \sum_{s=1}^{S} A_s \frac{d\mu^p_{t,t+s}}{dz_t^\top} + \sum_{s=1}^{S} B_s \frac{d\mu^d_{t,t+s}}{dz_t^\top} - \sum_{s=1}^{S} D_s \frac{dv^r_{t,t+s}}{dz_t^\top} + \frac{d\xi_t}{dz_t^\top}.$$

Assumption 1 sets the last term to zero. Therefore,

$$\frac{dq_t}{dz_t^\top} = A_0\Big(-\frac{dp_t}{dz_t^\top}\Big) + \sum_{s=1}^{S} A_s \frac{d\mu^p_{t,t+s}}{dz_t^\top} + \sum_{s=1}^{S} B_s \frac{d\mu^d_{t,t+s}}{dz_t^\top} - \sum_{s=1}^{S} D_s \frac{dv^r_{t,t+s}}{dz_t^\top}.$$

Premultiplying by a minus sign and postmultiplying by $\mathcal{M}_t(z_t)^{-1} = (dp_t/dz_t^\top)^{-1}$ as in (12) gives

$$\mathcal{E}_t(z_t) = A_0 - \sum_{s=1}^{S} A_s \frac{d\mu^p_{t,t+s}}{dz_t^\top}\left(\frac{dp_t}{dz_t^\top}\right)^{-1} - \sum_{s=1}^{S} B_s \frac{d\mu^d_{t,t+s}}{dz_t^\top}\left(\frac{dp_t}{dz_t^\top}\right)^{-1} + \sum_{s=1}^{S} D_s \frac{dv^r_{t,t+s}}{dz_t^\top}\left(\frac{dp_t}{dz_t^\top}\right)^{-1},$$

which is (15) by the definition of the shock-specific pass-through matrices. □

*Proof of Proposition 3.* Guess $p_t = \mathcal{M}\, z_t + \mathcal{N}\, \eta_t$. Under (18), $\mathbb{E}_t[z_{t+s}] = \phi^s\, z_t + \phi^{s-1}\, \eta_t$ for $s \geq 1$, and $\mathbb{E}_t[\eta_{t+k}] = 0$ for $k \geq 1$, so $\mathbb{E}_t[p_{t+s}] = \mathcal{M}\,\phi^s\, z_t + \mathcal{M}\,\phi^{s-1}\, \eta_t$. Substituting into (17) and collecting $z_t$ terms:

$$A_0\, \mathcal{M}\, z_t - \beta\, A_0\, \mathcal{M}\, \phi \sum_{s=0}^{\infty} (\delta\phi)^s\, z_t = \Lambda\, z_t.$$

The geometric series converges because $\delta\phi < (\beta+\delta)\phi < 1$. Summing:

$$A_0\, \mathcal{M} \cdot \frac{1-(\beta+\delta)\phi}{1-\delta\phi} = \Lambda.$$

Solving for $\mathcal{M}$ gives the multiplier in (19). Collecting $\eta_t$ terms:

$$A_0\, \mathcal{N} - \beta\, A_0\, \mathcal{M} \sum_{s=0}^{\infty} (\delta\phi)^s = 0, \qquad \text{i.e.,} \qquad \mathcal{N} = \frac{\beta}{1-\delta\phi}\, \mathcal{M}.$$

The condition $(\beta+\delta)\phi < 1$ ensures the scalar amplification factor is positive and finite. □

*Proof of Proposition 6.* The closed-form expression

$$\mathcal{M}_{jj} = \alpha(\beta,\delta,\phi)\, \left[(A_0)^{-1}\Lambda\right]_{jj}$$

is immediate from (19).

For the risk-pass-through statement, define

$$\mathcal{E}_t^{(0)} \equiv A_0 - \sum_{s=1}^{S} A_s \Pi_{t,s}^{p} - \sum_{s=1}^{S} B_s \Pi_{t,s}^{d}, \qquad G_t^{v} \equiv - \sum_{s=1}^{S} D_s \Pi_{t,s}^{v},$$

so that (15) becomes

$$\mathcal{E}_t = \mathcal{E}_t^{(0)} - G_t^{v}.$$

Assume

$$G_t^{v} \geq 0 \text{ entrywise}, \qquad \mathcal{E}_t(\tau) \equiv \mathcal{E}_t^{(0)} - \tau G_t^{v} \text{ is a nonsingular } M\text{-matrix for every } \tau \in [0,1]. \quad \text{(A.1)}$$

The first condition is the formal version of risk aversion plus inverse price-risk comovement after aggregating across the $m$ covariance channels, and the second implies $\mathcal{E}_t(\tau)^{-1} \geq 0$ entrywise. To be clear, if investors are risk averse in the sense that they demand less when there are higher variances and covariances, then $D_s \geq 0$ element-wise under our sign convention, since demand loads on risk as $-\sum_s D_s v_{t,t+s}^{r}$. The condition that risk moves inversely with prices means $\Pi_{t,s}^{v} \leq 0$ element-wise. Together, these imply $D_s \Pi_{t,s}^{v} \leq 0$ element-wise for each $s$, and therefore $G_t^{v} = -\sum_s D_s \Pi_{t,s}^{v} \geq 0$ element-wise as well.

Let $\mathcal{M}_t(\tau) \equiv \Lambda\, \mathcal{E}_t(\tau)^{-1}$. Then

$$\frac{d}{d\tau} \mathcal{E}_t(\tau)^{-1} = \mathcal{E}_t(\tau)^{-1} G_t^{v} \mathcal{E}_t(\tau)^{-1},$$

$$\frac{d}{d\tau} \mathcal{M}_t(\tau) = \Lambda\, \mathcal{E}_t(\tau)^{-1} G_t^{v} \mathcal{E}_t(\tau)^{-1}.$$

By (A.1), the right-hand side is entrywise nonnegative, so $d\mathcal{M}_t(\tau)/d\tau \geq 0$ entrywise. Hence $\mathcal{M}_t(1) \geq \mathcal{M}_t(0)$: introducing the variance channel lowers the instrumented elasticity and raises the multiplier relative to the no-risk-pass-through benchmark.

In the mean–variance benchmark, $(A_0^{\mathrm{mv}})^{-1}\Lambda = \gamma\beta\, \mathrm{Diag}(\bar{P})^{-1}\Sigma$, so

$$\mathcal{M}_{jj}^{\mathrm{mv}} = \alpha(\beta,\delta,\phi)\, \gamma\beta\, \frac{\Sigma_{jj}}{\bar{P}_j},$$

which is increasing in the own return variance $\Sigma_{jj}$.

□

# B Mean–Variance Benchmark

This appendix verifies that the $(\beta,\delta)$ equilibrium in Section 4 reduces to the standard mean–variance benchmark when $\delta = 0$.

**Demand in levels.** Consider a representative mean–variance investor with constant risk aversion $\gamma > 0$ and perceived dollar covariance matrix $\Sigma \succ 0$. Exact demand in levels is

$$Q_t = \frac{1}{\gamma}\,\Sigma^{-1}\big(\mathbb{E}_t[P_{t+1} + \mathcal{D}_{t+1}] - R^f\,P_t\big), \tag{A.2}$$

where $R^f = \beta^{-1}$ is the gross risk-free rate.

**Log-linearization.** Log-linearizing (A.2) around $(\bar{Q},\bar{P},\bar{\mathcal{D}})$ gives

$$q_t = -A_0^{\mathrm{mv}}\,p_t + A_1^{\mathrm{mv}}\,\mathbb{E}_t[p_{t+1}] + B_1^{\mathrm{mv}}\,\mathbb{E}_t[d_{t+1}],$$

with

$$\begin{aligned}
A_0^{\mathrm{mv}} &= \mathrm{Diag}(\bar{Q})^{-1}\,\frac{R^f}{\gamma}\,\Sigma^{-1}\,\mathrm{Diag}(\bar{P}),\\
A_1^{\mathrm{mv}} &= \mathrm{Diag}(\bar{Q})^{-1}\,\frac{1}{\gamma}\,\Sigma^{-1}\,\mathrm{Diag}(\bar{P}) = \beta\,A_0^{\mathrm{mv}},\\
B_1^{\mathrm{mv}} &= \mathrm{Diag}(\bar{Q})^{-1}\,\frac{1}{\gamma}\,\Sigma^{-1}\,\mathrm{Diag}(\bar{\mathcal{D}}).
\end{aligned}$$

The mean–variance model is therefore the $(\beta,\delta)$ parameterization of Section 4 with $\delta = 0$ and $\beta = 1/R^f$. Setting $\delta = 0$ collapses the infinite sequence $A_s = \beta\,\delta^{s-1}\,A_0$ to a single nonzero coefficient $A_1 = \beta\,A_0$; demand looks one period ahead but no further. In the language of Section 2.3, this nests the myopic mean–variance row of Table 1: the investor responds to one-period-ahead expected prices and dividends but has no intertemporal hedging motive and faces no adjustment costs. All the equilibrium results of Section 4 (the multiplier, the instrumented elasticity, and the comparative statics) apply with $\delta = 0$.

**Static multiplier building block.** Inverting $A_0^{\text{mv}}$ and multiplying by $\Lambda = \text{Diag}(\bar{Q})^{-1}$:

$$(A_0^{\text{mv}})^{-1} = \text{Diag}(\bar{P})^{-1} \frac{\gamma}{R^f} \Sigma \, \text{Diag}(\bar{Q}) = \gamma\beta \, \text{Diag}(\bar{P})^{-1} \Sigma \, \text{Diag}(\bar{Q}),$$
$$(A_0^{\text{mv}})^{-1}\Lambda = \gamma\beta \, \text{Diag}(\bar{P})^{-1} \Sigma \, \text{Diag}(\bar{Q}) \, \text{Diag}(\bar{Q})^{-1} = \gamma\beta \, \text{Diag}(\bar{P})^{-1} \Sigma.$$

**Equilibrium.** Setting $\delta = 0$ in Proposition 3 gives

$$\mathcal{M}^{\text{mv}} = \frac{1}{1 - \beta\phi} (A_0^{\text{mv}})^{-1}\Lambda = \frac{\gamma\beta}{1 - \beta\phi} \text{Diag}(\bar{P})^{-1}\Sigma.$$

The investor-level IE is

$$\mathcal{E}^{\text{mv}} = (1 - \beta\phi) \, A_0^{\text{mv}}.$$

The comparative statics in Section 4.5 specialize as follows. The own-price multiplier is $\mathcal{M}_{jj}^{\text{mv}} = \frac{\gamma\beta}{1-\beta\phi} \frac{\Sigma_{jj}}{\bar{P}_j}$, confirming that Proposition 6 yields $\mathcal{M}_{jj} \propto \Sigma_{jj}$ in this case. The amplification factor $\alpha = \frac{1}{1-\beta\phi}$ is increasing in $\phi$ (Proposition 4), and Proposition 5 is vacuous since $\delta$ is fixed at zero. For Proposition 7, the eigenvalues of $(A_0^{\text{mv}})^{-1}\Lambda = \gamma\beta \, \text{Diag}(\bar{P})^{-1}\Sigma$ coincide with those of the return covariance matrix (after price normalization), so flows along systematic risk directions command larger multipliers. The surprise-versus-expected decomposition (Proposition 8) specializes with $\delta = 0$: the attenuation factor uses $\beta$ in place of $\frac{\beta}{1-\delta\phi}$, so that $B_{\text{exp}} = \mathcal{M} \cdot [W + (1-\beta)\Sigma_\eta] \cdot [W + \Sigma_\eta]^{-1}$.

# C Integrability and Demand Coefficients

The integrability theorem (Proposition 3.H.1 in Mas-Colell et al., 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 $-D_s$ 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 $\Sigma$, then $A_0$ inherits the symmetry and positive definiteness of $\Sigma^{-1}$. 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 $\Sigma$ is symmetric and positive definite, so is $\Sigma^{-1}$. 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 derived in Appendix B:

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

where $\Lambda \equiv \text{Diag}(\bar{Q})^{-1}$ and $\beta = 1/R^f$. All higher-order coefficients vanish: $A_s = B_s = D_s = 0$ 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 $\tilde{A}_0 \equiv \Lambda^{-1} A_0 \,\text{Diag}(\bar{P})^{-1} = (R^f/\gamma)\,\Sigma^{-1}$ must be symmetric. This is the Slutsky symmetry condition expressed in the log-linear coefficients, and it holds because $\Sigma^{-1}$ 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

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

which identifies the composite object $\gamma^{-1}\Sigma^{-1}$. If the return covariance $\Sigma$ is separately observable from return data, inverting gives the risk-aversion parameter $\gamma$ directly. One may also use $B_1$ to form a second expression, $\gamma^{-1}\Sigma^{-1} = \Lambda^{-1} B_1 \,\text{Diag}(\bar{\mathcal{D}})^{-1}$, 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.

# D Sequential and Recursive Equilibrium Representations

This appendix expands on the discussion in Section 2.4 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, the equilibrium condition (8) directly links the structural demand coefficients $\{A_s\}$ to the impulse responses of equilibrium prices through the convolution equation (11). This equation must hold at every horizon $k$ 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 $\lambda$; 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 $\{A_s, B_s, D_s\}$ 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 (11) 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 $\{A_s, B_s, D_s\}$ 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.

# E Stacked Recovery System

For a given horizon $h$, equation (26) is linear in the unknown coefficient matrices. Applying $\mathrm{vec}(AB) = (B^\top \otimes I)\,\mathrm{vec}(A)$ gives

$$\begin{aligned}\mathrm{vec}(\Lambda\Psi_h^z) = (\Upsilon_h^\top \otimes I)\,\mathrm{vec}(A_0) - \sum_{s=1}^{S}(\Upsilon_{h+s}^\top \otimes I)\,\mathrm{vec}(A_s) \\ - \sum_{s=1}^{S}\left((\Phi_{s,h}^d)^\top \otimes I\right)\mathrm{vec}(B_s) - \sum_{s=1}^{S}\left((\Omega_{s,h})^\top \otimes I\right)\mathrm{vec}(D_s).\end{aligned} \tag{A.3}$$

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 vectorized coefficient matrices, but if all matrices are unrestricted, then $\theta = g(\theta)$, and thus the dimensionality of $\theta$ is simply the sum of the number of elements in all $A_s$, $B_s$, and $D_s$ matrices. Define

$$g_A(\theta) \equiv \left(\mathrm{vec}(A_0(\theta))^\top, \ldots, \mathrm{vec}(A_S(\theta))^\top\right)^\top, \qquad g_B(\theta) \equiv \left(\mathrm{vec}(B_1(\theta))^\top, \ldots, \mathrm{vec}(B_S(\theta))^\top\right)^\top,$$

$$g_D(\theta) \equiv \left(\mathrm{vec}(D_1(\theta))^\top, \ldots, \mathrm{vec}(D_S(\theta))^\top\right)^\top, \qquad g(\theta) \equiv \left(g_A(\theta)^\top, g_B(\theta)^\top, g_D(\theta)^\top\right)^\top.$$

Also define

$$m_H \equiv \left(\text{vec}(\Lambda\Psi_0^z)^\top, \ldots, \text{vec}(\Lambda\Psi_H^z)^\top\right)^\top, \qquad M_H \equiv \begin{bmatrix} M_0 \\ M_1 \\ \vdots \\ M_H \end{bmatrix},$$

where the horizon-$h$ block is

$$M_h \equiv \begin{bmatrix} M_h^A & M_h^B & M_h^D \end{bmatrix},$$

with

$$M_h^A \equiv \left[\Upsilon_h^\top \otimes I, \; -\Upsilon_{h+1}^\top \otimes I, \; \ldots, \; -\Upsilon_{h+S}^\top \otimes I\right],$$

$$M_h^B \equiv \left[-(\Phi_{1,h}^d)^\top \otimes I, \; \ldots, \; -(\Phi_{S,h}^d)^\top \otimes I\right], \qquad M_h^D \equiv \left[-(\Omega_{1,h})^\top \otimes I, \; \ldots, \; -(\Omega_{S,h})^\top \otimes I\right].$$

Stacking (A.3) across horizons $h = 0, \ldots, H$ yields equation (27) 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

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

where $D_s \in \mathbb{R}^{N \times m}$ 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 $B_s = D_s = 0$ 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 $B_s = D_s = 0$ 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 $F \ll N$.

# F Identification in Simple Cases

This appendix provides two simple examples that illustrate how the recovery system identifies structural parameters, how the formulas arise, and the economic intuition behind both the impulse responses and the associated rank condition.

Throughout, we consider a single stock and the one-step specification used in the main text. The recovery equation at horizon $h$ is

$$\Psi_h = C_1\left(\Xi_h - \beta\Xi_{h+1}\right) - D_1\Omega_{h+1}, \qquad h = 0,1.$$

Here, $\Psi_h$ is the flow (or market-clearing) impulse response at horizon $h$, $\Xi_h$ is the cumulative return impulse response, and $\Omega_{h+1}$ is the impulse response of future variance. The coefficient $C_1$ measures sensitivity to expected returns, while $D_1$ captures sensitivity to risk. Under the paper's sign convention, a positive $D_1$ means that higher future risk lowers demand.

## F.1 Horizon Intuition

The horizon $h$ indexes different points along the impulse-response path of a single shock. The $h = 0$ equation uses the impact response and its immediate continuation, while the $h = 1$ equation uses the next-period response after some of the shock has propagated. Persistent shocks therefore generate multiple distinct restrictions that must be satisfied by the same structural coefficients.

## F.2 Example 1: One Equation, One Unknown

Consider first the case with no variance term ($D_1 = 0$). The recovery equation at $h = 0$ reduces to

$$\Psi_0 = C_1\big(\Xi_0 - \beta\Xi_1\big),$$

so that

$$C_1 = \frac{\Psi_0}{\Xi_0 - \beta\Xi_1}. \tag{A.4}$$

**Intuition.** The numerator $\Psi_0$ is the observed quantity response—that is, how much of the outside order must be absorbed by the rest of the market. The denominator $\Xi_0 - \beta\Xi_1$ is the relevant expected-return wedge. This is essentially the pass-through of prices into expected returns, as discussed extensively by Davis, Kargar, and Li (2025). If the price decline is highly persistent (i.e., $\Xi_1 \approx \Xi_0$), then most of the price movement reflects a persistent repricing rather than a one-period return opportunity, and the denominator is small in magnitude. In that case, a large $C_1$ is required to rationalize the observed flow response.

Thus, equation (A.4) can be interpreted loosely as

$$C_1 = \frac{\text{quantity response}}{\text{expected-return incentive}}.$$

**Numerical Illustration.** To keep the arithmetic simple while staying close to the combined-shock recovery in the main text, choose values that imply a structural elasticity in the neighborhood of eight. Let

$$\Xi_0 = -0.030, \quad \Xi_1 = -0.028, \quad \beta = 0.99, \quad \Psi_0 = -0.0160.$$

Then

$$C_1 \approx \frac{-0.0160}{-0.030 - 0.99(-0.028)} \approx \frac{-0.0160}{-0.00228} \approx 7.0.$$

## F.3 Example 2: Two Equations, Two Unknowns

We now include the variance term. Using $h = 0,1$, the system becomes

$$\Psi_0 = C_1\big(\Xi_0 - \beta\Xi_1\big) - D_1\Omega_1,$$
$$\Psi_1 = C_1\big(\Xi_1 - \beta\Xi_2\big) - D_1\Omega_2.$$

Define

$$a_0 \equiv \Xi_0 - \beta\Xi_1, \qquad a_1 \equiv \Xi_1 - \beta\Xi_2.$$

Then the system can be written compactly as

$$\Psi_0 = a_0 C_1 - \Omega_1 D_1,$$
$$\Psi_1 = a_1 C_1 - \Omega_2 D_1.$$

Solving this $2 \times 2$ system yields

$$C_1 = \frac{\Psi_1 \Omega_1 - \Psi_0 \Omega_2}{a_1 \Omega_1 - a_0 \Omega_2}, \tag{A.5}$$

$$D_1 = \frac{a_0 \Psi_1 - a_1 \Psi_0}{a_1 \Omega_1 - a_0 \Omega_2}. \tag{A.6}$$

## F.4 Intuition for Numerators and Denominator

**Denominator (Rank Condition).** The denominator

$$a_1 \Omega_1 - a_0 \Omega_2$$

is the key identification object. It captures how the return wedge ($a_h$) and the risk impulse ($\Omega_h$) evolve differently across horizons.

If

$$\frac{a_0}{a_1} = \frac{\Omega_1}{\Omega_2},$$

then the return and risk channels decay in exactly the same proportion over time. In that case, the two equations are collinear, and the system is not full rank. Economically, this means that variation in flows cannot distinguish between return-driven and risk-driven demand responses.

Conversely, if

$$a_1 \Omega_1 - a_0 \Omega_2 \neq 0,$$

then the return and risk channels evolve differently across horizons. This generates independent variation, allowing $C_1$ and $D_1$ to be separately identified.

**Numerator for $C_1$.** The numerator in (A.5),

$$\Psi_1\Omega_1 - \Psi_0\Omega_2,$$

measures how the flow response co-moves with the risk path across horizons. It isolates the component of flows that cannot be explained purely by the risk channel, thereby identifying the return sensitivity $C_1$.

**Numerator for $D_1$.** The numerator in (A.6),

$$a_0\Psi_1 - a_1\Psi_0,$$

measures how flows evolve relative to the return wedge. It isolates the component of flows that cannot be explained purely by expected-return variation, thereby identifying the risk sensitivity $D_1$.

## F.5 Numerical Illustration

Let

$$\Xi_0 = -0.030, \quad \Xi_1 = -0.028, \quad \Xi_2 = -0.026, \quad \beta = 0.99,$$

so that

$$a_0 = -0.00228, \quad a_1 = -0.00226.$$

Let

$$\Psi_0 = -0.016064, \quad \Psi_1 = -0.018888,$$

and

$$\Omega_1 = -0.0004, \quad \Omega_2 = -0.0001.$$

Then

$$C_1 = 8.8, \qquad D_1 = 10.0.$$

**Interpretation.** Relative to the single-equation case, the inclusion of the variance channel implies that part of the observed response is attributable to changes in risk rather than expected returns. Here $D_1 > 0$ and $\Omega_h < 0$, so the risk term $-D_1\Omega_h$ is positive: worsening future risk attenuates

the amount of net buying generated by the return wedge alone. Because the variance response is front-loaded (i.e., $|\Omega_1| > |\Omega_2|$), the system attributes more of the impact-period response to the risk channel, which in turn requires a larger value of $C_1$ to jointly fit both horizons. In this illustration, controlling for variance raises the recovered elasticity from about 7.0 to 8.8, matching the qualitative pattern in the main recovery table.

### F.6 Summary

These examples illustrate that identification comes from differences in how impulse responses evolve over time. When return and risk channels exhibit distinct dynamic patterns, the recovery system can separately identify the corresponding structural coefficients. The full-rank condition requires precisely that these channels do not move in perfect proportion across horizons.

## G Additional Recovery Details and Results

This appendix writes the weekly recovery estimator in a unified minimum-distance language and makes explicit how each row of Table 8 is estimated. The estimator is a two-step procedure with generated moments. In the first step, we estimate the empirical impulse responses from the local projections in (32) and (33), and when risk is included we also estimate the weekly total-variance impulse responses defined in (30). In the second step, we choose the structural coefficients so that the model-implied flow impulse responses match the empirical flow impulse responses as closely as possible, conditional on the estimated return and variance responses.[13]

### G.1 General GMM setup

Let $z_{i,t}$ denote the combined weekly flow measure used in the main text. It is formed from FIT and OFI by a convex combination in their common shares-outstanding units, and the main specification uses a variance-weighted combination. For each horizon $h$, the first-step local projections deliver

[13] In the weekly implementation we fix the discount factor at $\beta = 0.999$. This is close to a standard weekly discount factor and we do not attempt to identify it separately from the continuation-shape parameter $\delta$. Treating $\beta$ as fixed keeps the low-dimensional recovery focused on the persistence of demand through $\delta$, rather than on small differences in the discount rate that the weekly impulse responses are unlikely to pin down precisely.

empirical diagonal and average-other impulse responses

$$\hat{\Psi}_h^{\text{diag}}, \quad \hat{\Psi}_h^{\text{off}}, \qquad \hat{\Xi}_h^{\text{diag}}, \quad \hat{\Xi}_h^{\text{off}},$$

and, in the specifications with risk,

$$\hat{\Omega}_h^{\text{diag}}, \quad \hat{\Omega}_h^{\text{off}}.$$

The objects $\hat{\Psi}$ are the empirical flow responses, $\hat{\Xi}$ are the cumulative return responses, and $\hat{\Omega}$ are the weekly total-variance responses. The second-step estimator treats the return and variance impulse responses as generated first-step inputs and chooses the structural coefficients to match the corresponding flow impulse responses implied by the model.

This appendix works with the same reduced system as the main text. It does not estimate the unrestricted $(H+1)N^2$ equations from Proposition 9. Instead, for each horizon it keeps two generated moments, an own response and an average-other response. Thus the second step is based on $2(H+1)$ flow equations. The coefficients $(A_0^{\text{diag}}, A_0^{\text{off}})$, $(C_s^{\text{diag}}, C_s^{\text{off}})$, and $(D_s^{\text{diag}}, D_s^{\text{off}})$ should therefore be interpreted as parameters of this reduced own/average-other system.

To write the second-step estimator, fix a specification and let $\theta$ denote its parameter vector. Collect the selected diagonal and average-other flow moments into the vector

$$\hat{y} \equiv \begin{pmatrix} \hat{\Psi}_0^{\text{diag}} \\ \vdots \\ \hat{\Psi}_H^{\text{diag}} \\ \hat{\Psi}_0^{\text{off}} \\ \vdots \\ \hat{\Psi}_H^{\text{off}} \end{pmatrix} \in \mathbb{R}^{2(H+1)},$$

where $H$ is the largest flow horizon used by that specification. Let

$$\hat{\mu}(\theta) \equiv \begin{pmatrix} \mu_0^{\text{diag}}(\theta) \\ \vdots \\ \mu_H^{\text{diag}}(\theta) \\ \mu_0^{\text{off}}(\theta) \\ \vdots \\ \mu_H^{\text{off}}(\theta) \end{pmatrix}$$

denote the vector of model-implied flow impulse responses, constructed from the empirical return and variance impulse responses together with the structural restrictions of the specification under consideration. The second-step moment vector is then

$$\hat{\varphi}(\theta) \equiv \hat{y} - \hat{\mu}(\theta)$$

The reported estimator solves

$$\hat{\theta} \in \arg\min_{\theta} \ \hat{\varphi}(\theta)^{\top} W \hat{\varphi}(\theta).$$

Because there is only one combined-flow system, the weighting matrix is a single diagonal matrix,

$$W = \omega I_{2(H+1)},$$

which is equivalent, up to scale, to ordinary least squares on the stacked own and average-other equations. In practice, the reported estimates are therefore the minimum-distance coefficients that best fit the observed own and average-other flow responses conditional on the estimated return and variance responses.

It is helpful to separate the linear and nonlinear cases. In the linear specifications, $\theta$ collects only coefficients that enter linearly. Writing the model-implied moments as $\hat{\mu}(\theta) = X\theta$, the estimator reduces to weighted least squares,

$$\hat{\theta} = \left(X^{\top} W X\right)^{-1} X^{\top} W \hat{y},$$

where $X$ and $\hat{y}$ stack the own and average-other equations across horizons. In the nonlinear cases, the only nonlinear parameter is the decay parameter $\delta$. Conditional on a candidate $\delta$, the model is

linear in the remaining coefficients. Thus the estimator can be written as

$$\hat{\alpha}(\delta) = \big(X(\delta)^\top W X(\delta)\big)^{-1} X(\delta)^\top W \hat{y},$$

followed by the one-dimensional outer optimization

$$\hat{\delta} \in \arg\min_{\delta} \hat{\varphi}\big(\hat{\alpha}(\delta),\delta\big)^\top W \hat{\varphi}\big(\hat{\alpha}(\delta),\delta\big).$$

Equivalently, one may view the procedure as searching over $\delta$ and, for each candidate, evaluating the corresponding weighted sum of squared moment errors.

Throughout, the superscript "diag" denotes the response of the shocked stock itself, while "off" denotes the average response across the other stocks.

## G.2 Standard errors

For every row of Table 8, standard errors are obtained by resampling weeks with replacement and re-estimating the full second-step recovery on the resampled first-step weekly coefficient series. In each bootstrap draw, we recompute the first-step impulse responses from the resampled weekly local-projection coefficients and then re-estimate the structural coefficients. The reported standard errors are the bootstrap standard deviations of the resulting parameter draws. This resampling procedure propagates both first-step estimation uncertainty in the impulse responses and second-step uncertainty in the structural recovery.