On the Recovery and Usage of Demand Elasticities in Dynamic Settings*

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

March 26, 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 under a weak exclusion restriction 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 4.3 to 6.2 far exceed the naive inverse-multiplier benchmark of 0.3 to 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 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 (Li, Sokolinski, and Tamoni, 2026). 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 can violate the exclusion restriction to 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 and a weak price-mediated exclusion restriction, the inverse of the instrumented elasticity is the price multiplier for the identifying shock. The exclusion restriction requires only that expected dividends, future prices, and risk enter demand through the equilibrium price path induced by the shock, not through a direct channel from the instrument to demand. Stronger exogeneity arguments (for example, that the identifying flow is unrelated to cash-flow news) can sharpen the interpretation in applications, but the baseline restriction is all that is needed.

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 0.3 to 0.4. Once we use the dynamic path of price and flow responses to recover structural demand, the implied own-price elasticity rises to about 4.3 to 6.2, depending on the restriction imposed on the forward-looking coefficients. 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 develops the unified demand system and defines the structural and instrumented elasticities. Section 3 connects the instrumented elasticity to the price multiplier. Section 4 derives comparative statics in a benchmark CARA equilibrium. Section 5 tests these predictions empirically. Section 6 develops the recovery theory and reports the empirical recovery results. Section 7 concludes.

2 Different Types of Price Elasticities

This section establishes the theoretical foundation for the rest of the paper. We begin with the Campbell-Shiller identity, which imposes an accounting constraint: any price movement must pass through to expected dividends, expected returns, or both. Because different assumptions about what adjusts when prices move produce different elasticities even under the same underlying demand system, there is no single price elasticity in dynamic asset markets. We then develop a unified log-linear demand framework that nests the main specifications used in the DSAP literature, link the structural demand coefficients through the Campbell-Shiller restrictions, and use concrete examples to show that commonly studied elasticities (static, variance-based, and dynamic) are distinct objects corresponding to different pass-through patterns.

2.1 Campbell-Shiller approximation and price pass-through

Let $Q_t \in \mathbb{R}^N$ denote portfolio holdings, $P_t \in \mathbb{R}^N$ prices, and $D_t \in \mathbb{R}^N$ dividends. Define lower-case variables as logs, so that $p_t \in \mathbb{R}^N$ and $d_t \in \mathbb{R}^N$ denote log prices and log dividends. The one-period Campbell-Shiller approximation is

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

where $\rho \in (0,1)$ and κ is the log-linearization constant.

Iterating forward under the usual transversality condition yields the long-run representation

$$p_t = \bar{\kappa} + (1-\rho) \sum_{j=1}^{\infty} \rho^{j-1} \mathbb{E}_t[d_{t+j}] - \sum_{j=1}^{\infty} \rho^{j-1} \mathbb{E}_t[r_{t+j}], \tag{2}$$

horizon j	$\frac{I}{(1-\rho)\rho^{j-1}} =$	$\frac{\partial \mathbb{E}_t[d_{t+j}]}{\partial p_t^\top}$	$\frac{I}{\rho^{j-1}} =$	$\left(-\frac{\partial \mathbb{E}_t[r_{t+j}]}{\partial p_t^\top}\right)$
1	$\left[(1-\rho)\rho^0\right]^{-1}$	27.6243	ρ^{-0}	1.0000
2	$\left[(1-\rho)\rho^1\right]^{-1}$	28.6619	ρ^{-1}	1.0376
3	$\left[(1-\rho)\rho^2\right]^{-1}$	29.7384	ρ^{-2}	1.0765
4	$\left[(1-\rho)\rho^3\right]^{-1}$	30.8554	ρ^{-3}	1.1170
5	$\left[(1-\rho)\rho^4\right]^{-1}$	32.0143	ρ^{-4}	1.1589

Table 1. Pass-throughs.
Diagonal pass-through magnitudes implied by Campbell-Shiller under $\rho = 0.9638$. The left pair of columns reports the mapping when a price change is absorbed entirely through expected dividends at horizon j; the right pair reports the mapping when it is absorbed entirely through expected returns at horizon j.

where $\bar{\kappa}$ is the implied constant. Differentiating (2) with respect to p_t gives

$$I = (1-\rho)\sum_{j=1}^{\infty}\rho^{j-1}\frac{\partial \mathbb{E}_t[d_{t+j}]}{\partial p_t^\top} - \sum_{j=1}^{\infty}\rho^{j-1}\frac{\partial \mathbb{E}_t[r_{t+j}]}{\partial p_t^\top}. \tag{3}$$

Equation (3) is the key accounting restriction in the paper. A price movement cannot occur in isolation. It must pass through to expected dividends, expected returns, or both. Any elasticity concept in finance therefore has to specify that pass-through pattern.

Table 1 gives a simple calibration of these pass-through magnitudes using annual $\rho = 0.9638$ from Cochrane (2008). If a price movement is absorbed entirely through expected dividends at horizon j, the diagonal pass-through is $\left[(1-\rho)\rho^{j-1}\right]^{-1}$. If it is instead absorbed entirely through expected returns at horizon j, the diagonal pass-through is $\rho^{-(j-1)}$. The table makes clear that longer-horizon pass-through requires larger revisions in the corresponding expected dividend or expected return term to rationalize the same price movement today.

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

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

and the conditional second moments

$$V^p_{t,t+s} \equiv \mathbb{V}\mathrm{ar}_t(p_{t+s}), \qquad V^d_{t,t+s} \equiv \mathbb{V}\mathrm{ar}_t(d_{t+s}).$$

Let

$$m \equiv \frac{N(N+1)}{2}, \qquad v^p_{t,t+s} \equiv \mathrm{vech}(V^p_{t,t+s}) \in \mathbb{R}^m, \qquad v^d_{t,t+s} \equiv \mathrm{vech}(V^d_{t,t+s}) \in \mathbb{R}^m,$$

Thus the second-moment state variables are vectors of unique covariance terms rather than $N \times N$ matrices.

Equation (1) implies the corresponding conditional mean relation

$$\mathbb{E}_t[r_{t+1}] = \kappa + \rho\mu^p_{t,t+1} + (1-\rho)\mu^d_{t,t+1} - p_t.$$

It also implies that the conditional variance of the Campbell-Shiller continuation term can be decomposed as

$$\mathbb{V}\mathrm{ar}_t\left(\rho p_{t+s} + (1-\rho)d_{t+s}\right) = \rho^2 V^p_{t,t+s} + (1-\rho)^2 V^d_{t,t+s} + 2\rho(1-\rho)\,\mathrm{Sym}\left(\Gamma^{pd}_{t,t+s}\right),$$

where

$$\Gamma^{pd}_{t,t+s} \equiv \mathrm{Cov}_t(p_{t+s}, d_{t+s}), \qquad \mathrm{Sym}(M) \equiv \frac{1}{2}(M + M^\top).$$

Thus the two covariance terms enter only through the symmetric part of the price-dividend

covariance matrix. Define

$$V^x_{t,t+s} \equiv \mathbb{V}\mathrm{ar}_t\big(\rho p_{t+s} + (1-\rho) d_{t+s}\big), \qquad v^x_{t,t+s} \equiv \mathrm{vech}(V^x_{t,t+s}) \in \mathbb{R}^m.$$

For $s = 1$, the current price, p_t, drops out of the return variance because it is known at time t. For $s \geq 2$, we work with the continuation-value variance $V^x_{t,t+s}$ rather than the exact variance of r_{t+s}, which would additionally require lagged-price and cross-horizon covariance terms.

2.2 A one-period demand example

Before introducing the full intertemporal framework, it is useful to start with the simplest case in which investors care only about next-period expected returns and next-period return risk. Consider local demand of the form

$$q_t = \bar{q} + C_1 \mathbb{E}_t[r_{t+1}] + \tilde{C}_1 \,\mathrm{vech}\left(\mathbb{V}\mathrm{ar}_t(r_{t+1})\right). \tag{4}$$

Because p_t is known at time t, the one-period Campbell-Shiller identity implies

$$\mathbb{E}_t[r_{t+1}] = \kappa + \rho \mu^p_{t,t+1} + (1-\rho)\mu^d_{t,t+1} - p_t. \tag{5}$$

For the covariance terms, define

$$v^{pd}_{t,t+1} \equiv \mathrm{vech}\left(\mathrm{Cov}_t(p_{t+1}, d_{t+1})\right).$$

Since $\mathrm{Cov}_t(d_{t+1},p_{t+1}) = \mathrm{Cov}_t(p_{t+1},d_{t+1})^\top$, this single object already contains the same information as the reverse-order covariance. Then

$$\mathbb{V}\mathrm{ar}_t(r_{t+1}) = \rho^2 V^p_{t,t+1} + (1-\rho)^2 V^d_{t,t+1} + 2\rho(1-\rho)\,\mathrm{Sym}\left(\mathrm{Cov}_t(p_{t+1},d_{t+1})\right). \tag{6}$$

Substituting (5) and (6) into (4) yields

$$\begin{aligned} q_t = \bar{q} - C_1 p_t + \rho C_1 \mu^p_{t,t+1} + (1-\rho) C_1 \mu^d_{t,t+1} \\ + \tilde{C}_1\,\mathrm{vech}\left(\rho^2 V^p_{t,t+1} + (1-\rho)^2 V^d_{t,t+1} + 2\rho(1-\rho)\,\mathrm{Sym}\left(\mathrm{Cov}_t(p_{t+1},d_{t+1})\right)\right). \end{aligned}$$

This one-period example already contains the core logic of the paper. If demand is written over returns, then current prices matter only because they help determine expected returns, while return risk can be decomposed into price-risk, dividend-risk, and price-dividend covariance terms.

2.3 A unified local demand system

Fix a reference point $(\bar{Q},\bar{P})$ with strictly positive components and define log deviations

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

The main specification we work with treats demand as a function of the expected path of future returns and continuation risk:

$$q_t = \bar{q} + \sum_{s=1}^{S} C_s \mathbb{E}_t[r_{t+s}] + \sum_{s=1}^{S} \tilde{C}_s v^x_{t,t+s}, \tag{7}$$

where $C_s \in \mathbb{R}^{N\times N}$ and $\tilde{C}_s \in \mathbb{R}^{N\times m}$ in the multivariate case, with $m = N(N+1)/2$.[2] We take the C_s coefficients to be the primary objects in the paper. They describe how current holdings respond to expected returns at each future horizon.

Using (5), (7) can be written in the equivalent price-dividend form

$$q_t = -A_0 p_t + \sum_{s=1}^{S} A_s \mu^p_{t,t+s} + \sum_{s=1}^{S} B_s \mu^d_{t,t+s} + \sum_{s=1}^{S} \tilde{C}_s v^x_{t,t+s}, \tag{8}$$

where

$$A_0 = C_1, \qquad A_s = \rho C_s - C_{s+1} \text{ for } s = 1,\ldots,S, \qquad B_s = (1-\rho)C_s \text{ for } s = 1,\ldots,S, \tag{9}$$

with the convention $C_{S+1} \equiv 0$. Equation (8) is the price-dividend representation implied by the return-based system. One could allow unrestricted matrices A_s and B_s that differ from the C-implied restrictions, but economically that amounts to allowing direct sensitivity to future prices or dividends beyond their role in expected returns. We do not rule out that extension, but the paper treats the C_s system as the main object.

For most of the theoretical analysis below, we hold conditional second moments and expected dividends fixed and absorb them into a reduced-form shifter:

$$q_t = -A_0 p_t + \sum_{s=1}^{S} A_s \mu^p_{t,t+s} + f_t, \tag{10}$$

where

$$f_t \equiv \sum_{s=1}^{S} B_s \mu^d_{t,t+s} + \sum_{s=1}^{S} \tilde{C}_s v^x_{t,t+s}.$$

[2]For scalar notation, the coefficients collapse to numbers. We retain the symbol $\tilde{C}_s$ for the continuation-risk coefficient even though its number of columns differs from that of C_s.

This specialization is not a restriction on the notation. It only says that the benchmark equilibrium and recovery arguments focus on the price channel and treat the remaining state variables as part of a composite demand shifter.

2.4 Decomposing demand into price and dividend terms

For horizons $s = 1, \ldots, S$, the Campbell-Shiller identity implies

$$\mathbb{E}_t[r_{t+s}] = \kappa + \rho\mu^p_{t,t+s} + (1-\rho)\mu^d_{t,t+s} - \mu^p_{t,t+s-1},$$

with the convention $\mu^p_{t,t} \equiv p_t$. The coefficient restrictions in (9) therefore show exactly how the price-dividend coefficients are generated by the return coefficients. In this sense, the A_s and B_s matrices are derived objects in the main specification, while the C_s matrices are primitive.

At first glance, the contemporaneous coefficient 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. Equation (3) shows that 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 coherent elasticities become relevant: one can ask about a price movement that maps into next-period discount rates, longer-horizon discount rates, or expected cash flows, and each case produces a distinct elasticity even under the same underlying demand system.

2.5 Examples of coherent elasticities

The log-linear specification is deliberately flexible: it allows several coherent notions of an elasticity, each corresponding to a different pass-through pattern in (3). To make that explicit, define the pass-through matrices

$$\Pi_s^r \equiv -\frac{\partial \mathbb{E}_t[r_{t+s}]}{\partial p_t^\top}, \qquad \Pi_s^d \equiv \frac{\partial \mathbb{E}_t[d_{t+s}]}{\partial p_t^\top}.$$

Under the return-based system (7), the elasticity associated with a generic pass-through pattern is

$$-\frac{\partial q_t}{\partial p_t^\top} = \sum_{s=1}^{S} C_s \Pi_s^r. \tag{11}$$

If one instead allows the unrestricted price-dividend generalization (8), then

$$-\frac{\partial q_t}{\partial p_t^\top} = \sum_{s=1}^{S} C_s \Pi_s^r - \sum_{s=1}^{S} B_s \Pi_s^d, \tag{12}$$

where the second term vanishes in the main return-based specification.

One-step discount-rate elasticity. Suppose the price movement is absorbed entirely by next-period expected returns, so that $\Pi_1^r = I$ and $\Pi_s^r = 0$ for $s \geq 2$, while $\Pi_s^d = 0$ for all s. Then (11) gives

$$-\frac{\partial q_t}{\partial p_t^\top} = C_1.$$

This is the sensitivity to a one-percent decline in prices that maps one-for-one into a one-percent increase in next-period expected returns.

Longer-horizon discount-rate elasticity. Suppose instead that the price movement is absorbed entirely by expected returns at horizon $h \geq 2$. Then (3) implies

$$\Pi_h^r = \rho^{-(h-1)} I, \qquad \Pi_s^r = 0 \text{ for } s \neq h, \qquad \Pi_s^d = 0 \text{ for all } s.$$

Hence

$$-\frac{\partial q_t}{\partial p_t^\top} = \rho^{-(h-1)} C_h.$$

For $h = 2$, this becomes $\rho^{-1} C_2$; for $h = 3$, it becomes $\rho^{-2} C_3$; and so on. If demand is myopic, then $C_h = 0$ for all $h \geq 2$, so all longer-horizon discount-rate elasticities are zero by construction.

Cash-flow elasticity. Suppose the price movement is absorbed entirely by expected dividends at horizon h, with all expected returns fixed. Then (3) implies

$$\Pi_h^d = \left[(1-\rho)\rho^{h-1}\right]^{-1} I, \qquad \Pi_s^d = 0 \text{ for } s \neq h, \qquad \Pi_s^r = 0 \text{ for all } s.$$

Under the main return-based system, (11) implies

$$-\frac{\partial q_t}{\partial p_t^\top} = 0.$$

If one instead allows the unrestricted dividend coefficients B_h, then (12) yields

$$-\frac{\partial q_t}{\partial p_t^\top} = -\left[(1-\rho)\rho^{h-1}\right]^{-1} B_h.$$

This is precisely why we view unrestricted B_h terms as a departure from the main return-based specification: they load directly on cash-flow news rather than only on the returns that cash-flow news implies.

2.6 There is no "single" price elasticity in finance

These examples make clear why searching for "the" elasticity is misguided in dynamic asset markets. Different empirical designs generate different pass-through patterns across horizons and across the dividend versus return channels, and therefore recover distinct objects even when all are consistent with the same underlying structural demand system. The right question is not which elasticity is the true one, but which is relevant for the counterfactual at hand.

This point is especially important for empirical work. Much of the demand-system 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 instrumented elasticities that blend structural demand slopes with the propagation of the instrument through the term structure of expected returns.

A useful perspective is that asset demand is indirect. Investors do not value prices per se. They value 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.

3 Instrumented demand elasticities

Which of the many coherent elasticities from Section 2 does an empiricist actually recover? The answer depends on the instrument. A shock that moves prices also reshapes the expected path of future returns and dividends, so the measured elasticity reflects the particular pass through pattern that the instrument induces. This section makes that logic precise. We define the instrumented elasticity (IE) as the reduced form demand response divided by the first stage price response, introduce the price multiplier matrix, and ask when the two coincide. Proposition 1 provides the answer: under linearized market clearing and an exclusion restriction we call *price-mediated exogeneity*, which requires that the instrument affects demand only through the equilibrium price path, the aggregate IE is the inverse of the price multiplier. We close the section by decomposing the IE into the structural demand coefficients C_s and characterizing the special cases that arise when the instrument operates through cash flows, through variance, through a single horizon of expected returns, or through the full term structure.

3.1 Set-up and definition

Recall the demand specification (10), where conditional second moments and expected dividends have been absorbed into the reduced form shifter f_t. That equation draws a sharp line between two objects that the literature often conflates.

Structural demand elasticity. Holding fixed all future expected prices $\mu^p_{t,t+s}$ and the 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.

Instrumented elasticity matrix. Suppose now that prices are moved by an observable shock z_t that also changes the expected future path of prices. Then the IE matrix, $\mathcal{E}_t(z_t)$, is defined as

$$\mathcal{E}_t(z_t) \equiv \underbrace{-\frac{dq_t}{dz_t^\top}}_{\text{reduced form}} \overbrace{\left(\frac{dp_t}{dz_t^\top}\right)^{-1}}^{\text{first stage}}. \tag{13}$$

Price multiplier matrix. The first stage in (13) defines the price multiplier matrix:

$$\mathcal{M}_t(z_t) \equiv \frac{dp_t}{dz_t^\top}, \tag{14}$$

so that the IE can equivalently be written as

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

3.2 Linearized market clearing

Let outside holdings be $Z_t \in \mathbb{R}^N$, with reference point $\bar{Z}$, and suppose total shares are fixed at $\mathbf{1}_N$ so that

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

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

$$\mathrm{Diag}(\bar{Q})\, q_t + z_t = 0, \qquad z_t \equiv Z_t - \bar{Z}. \tag{15}$$

Define $\Lambda \equiv \text{Diag}(\bar{Q})^{-1}$. Then (15) implies

$$q_t = -\Lambda z_t. \tag{16}$$

Substituting (16) into (10) gives the linear equilibrium condition

$$A_0 p_t - \sum_{s=1}^{S} A_s \mathbb{E}_t[p_{t+s}] = \Lambda z_t + f_t. \tag{17}$$

3.3 Indirect and direct characteristic effects

Stack the demand-relevant characteristics as

$$\mathcal{X}_t \equiv \left(\mu^p_{t,t+1}, \ldots, \mu^p_{t,t+S}, \mu^d_{t,t+1}, \ldots, \mu^d_{t,t+S}, v^x_{t,t+1}, \ldots, v^x_{t,t+S}\right).$$

For each horizon s, define the non-price-mediated responses

$$\begin{aligned}
\Delta^p_{t,s}(z_t) &\equiv \frac{d\mu^p_{t,t+s}}{dz_t^\top} - \left.\frac{\partial \mu^p_{t,t+s}}{\partial p_t^\top}\right|_{z_t} \frac{dp_t}{dz_t^\top}, \\
\Delta^d_{t,s}(z_t) &\equiv \frac{d\mu^d_{t,t+s}}{dz_t^\top} - \left.\frac{\partial \mu^d_{t,t+s}}{\partial p_t^\top}\right|_{z_t} \frac{dp_t}{dz_t^\top}, \\
\Delta^v_{t,s}(z_t) &\equiv \frac{dv^x_{t,t+s}}{dz_t^\top} - \left.\frac{\partial v^x_{t,t+s}}{\partial p_t^\top}\right|_{z_t} \frac{dp_t}{dz_t^\top}.
\end{aligned}$$

Total differentiation of (8) then gives

$$\frac{dq_t}{dz_t^\top} = \underbrace{\left[-A_0 + \sum_{s=1}^{S} A_s \left.\frac{\partial \mu^p_{t,t+s}}{\partial p_t^\top}\right|_{z_t} + \sum_{s=1}^{S} B_s \left.\frac{\partial \mu^d_{t,t+s}}{\partial p_t^\top}\right|_{z_t} + \sum_{s=1}^{S} \tilde{C}_s \left.\frac{\partial v^x_{t,t+s}}{\partial p_t^\top}\right|_{z_t}\right] \frac{dp_t}{dz_t^\top}}_{\text{indirect effect through price-mediated characteristics}}$$
$$+ \underbrace{\sum_{s=1}^{S} A_s \Delta^p_{t,s}(z_t) + \sum_{s=1}^{S} B_s \Delta^d_{t,s}(z_t) + \sum_{s=1}^{S} \tilde{C}_s \Delta^v_{t,s}(z_t)}_{\text{direct characteristic effect}} . \tag{18}$$

The decomposition makes explicit what an exclusion restriction must rule out. The instrument may well move expected future prices, expected dividends, future price variance, future dividend variance, and the price-dividend covariance terms embedded in $v^x_{t,t+s}$. The question is whether those responses are entirely induced by the contemporaneous price movement or whether the shock has an independent direct effect on characteristics.

3.4 Price-mediated exogeneity

The exclusion restriction we impose is that the direct characteristic effect in (18) vanishes:

$$\Delta^p_{t,s}(z_t) = 0, \qquad \Delta^d_{t,s}(z_t) = 0, \qquad \Delta^v_{t,s}(z_t) = 0 \qquad \text{for all } s = 1, \ldots, S. \tag{19}$$

Under (19), the instrument affects demand only through the equilibrium price path; that is, all of its effects are "price-mediated". The shock may still change expected future prices, dividends, and second moments, but only because current prices change first.

Proposition 1 (Instrumented elasticity and the multiplier matrix)**.** *Suppose:*

1. *the price multiplier matrix $\mathcal{M}_t(z_t)$ in (14) is invertible;*

2. *linearized market clearing* (16) *holds.*

Then

$$\mathcal{E}_t(z_t) = \Lambda \mathcal{M}_t(z_t)^{-1}. \tag{20}$$

Define the aggregate demand elasticity matrix by

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

Then

$$\tilde{\mathcal{E}}_t(z_t) = \mathcal{M}_t(z_t)^{-1}. \tag{21}$$

The aggregate elasticity $\tilde{\mathcal{E}}_t(z_t)$ rescales that slope into shares outstanding units, matching the units of the instrument z_t. In those units the aggregate elasticity is the inverse of the price multiplier.[3]

It is worth stressing what price mediated exogeneity in (19) does and does not require. The instrument need not leave future prices, dividends, or second moments unchanged; it may move all of them. What is ruled out is a channel through which z_t shifts those objects, or demand itself, independently of the contemporaneous price response.

3.5 What the IE loads on

Under price-mediated exogeneity the direct term in (18) vanishes, and dividing by the first stage gives

$$\mathcal{E}_t(z_t) = A_0 - \sum_{s=1}^{S} A_s \Pi^p_{t,s}(z_t) - \sum_{s=1}^{S} B_s \Pi^d_{t,s}(z_t) - \sum_{s=1}^{S} \tilde{C}_s \Pi^v_{t,s}(z_t), \tag{22}$$

[3]With multiple elastic investors, 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. In the one investor benchmark this reduces to $\tilde{\mathcal{E}}_t(z_t) = \text{Diag}(\bar{Q})\mathcal{E}_t(z_t)$.

where the pass-throughs are defined as

$$\Pi^{p}_{t,s}(z_t) \equiv \left.\frac{\partial \mathbb{E}_t[p_{t+s}]}{\partial p_t^\top}\right|_{z_t}, \qquad \Pi^{d}_{t,s}(z_t) \equiv \left.\frac{\partial \mathbb{E}_t[d_{t+s}]}{\partial p_t^\top}\right|_{z_t}, \qquad \Pi^{v}_{t,s}(z_t) \equiv \left.\frac{\partial v^{x}_{t,t+s}}{\partial p_t^\top}\right|_{z_t}.$$

The IE therefore loads on three channels: pass through to future expected prices, to future expected dividends, and to future continuation risk $v^x_{t,t+s}$ (which itself embeds future price variance, future dividend variance, and price dividend covariance terms).

In the return-based specification, it is often more transparent to work with return pass-through. Define

$$\Pi^{r}_{t,s}(z_t) \equiv -\left.\frac{\partial \mathbb{E}_t[r_{t+s}]}{\partial p_t^\top}\right|_{z_t}.$$

Then

$$\mathcal{E}_t(z_t) = \sum_{s=1}^{S} C_s \Pi^{r}_{t,s}(z_t) - \sum_{s=1}^{S} \tilde{C}_s \Pi^{v}_{t,s}(z_t).$$

The first term captures the indirect effect through expected returns at each horizon. The second captures the effect of risk revaluation.

The first sum captures how the instrument moves expected returns at each horizon; the second captures how it revalues risk.

3.6 Special cases: Static, variance, immediate, and dynamic IEs

Only after imposing stronger restrictions does the IE isolate narrower structural objects.

Cash flow IE. Suppose the instrument moves prices only through expected dividends, with no expected return or variance effects:

$$\Pi^{r}_{t,s}(z_t) = 0, \qquad \Pi^{v}_{t,s}(z_t) = 0 \qquad \text{for all } s.$$

Under the main return-based specification,

$$\mathcal{E}_t^{\mathrm{cf}}(z_t) = 0.$$

With unrestricted dividend coefficients, the relevant object is the cash flow component of (22).

Variance IE. Suppose the instrument changes conditional second moments while leaving expected returns unchanged:

$$\Pi_{t,s}^{r}(z_t) = 0 \qquad \text{for all } s.$$

Then

$$\mathcal{E}_t^{\mathrm{var}}(z_t) = -\sum_{s=1}^{S} \tilde{C}_s \Pi_{t,s}^{v}(z_t).$$

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

Immediate return IE. Suppose the instrument generates a pure one-period discount-rate movement with no variance effects:

$$\Pi_{t,1}^{r}(z_t) = I, \qquad \Pi_{t,s}^{r}(z_t) = 0 \text{ for all } s \geq 2, \qquad \Pi_{t,s}^{v}(z_t) = 0 \text{ for all } s.$$

Then

$$\mathcal{E}_t^{\mathrm{imm}}(z_t) = C_1.$$

Horizon-h return IE. If the instrument only affects horizon h and variance is unchanged, then

$$\mathcal{E}_t^{(h)}(z_t) = C_h \Pi_{t,h}^{r}(z_t).$$

For the pure Campbell-Shiller horizon-h discount-rate experiment, $\Pi^r_{t,h}(z_t) = \rho^{-(h-1)} I$, so

$$\mathcal{E}^{(h)}_t(z_t) = \rho^{-(h-1)} C_h.$$

Dynamic return IE. If the instrument shifts the whole expected-return term structure and possibly continuation risk, then

$$\mathcal{E}^{\text{dyn}}_t(z_t) = \sum_{s=1}^{S} C_s \Pi^r_{t,s}(z_t) - \sum_{s=1}^{S} \tilde{C}_s \Pi^v_{t,s}(z_t). \tag{23}$$

This is the empirically relevant case for persistent flow instruments. Such designs generally identify neither A_0 nor a single C_s; they identify the weighted combination in (23), with weights given by the pass-through of the shock into the term structure of expected returns and continuation risk.

This is the empirically relevant case. Persistent flow instruments generally identify neither A_0 nor any single C_s. They identify the weighted combination in (23), with weights determined by how the shock propagates into expected returns and risk at each horizon.

Once the stochastic process for (z_t, f_t) is specified, (17) pins down the price process. Section 4 works this out for a benchmark CARA model.

4 Elasticity as a Function of Key Variables

Up to this point, the paper has emphasized that dynamic environments admit multiple elasticity concepts and that instrumented elasticities generally differ from structural derivatives. This section shows that the resulting objects are still tractable and economically interpretable in equilibrium. We study a benchmark mean-variance model—which can arise from CARA set ups, or CRRA set ups via an approximation, rewrite it in the unified log-linear notation, and then derive 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 Benchmark demand

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

$$Q_t = \frac{1}{\gamma}\Sigma^{-1}\left(\mathbb{E}_t[P_{t+1} + D_{t+1}] - R^f P_t\right), \tag{24}$$

where $R^f = \beta^{-1}$ is the gross risk-free rate determined exogenously. Log-linearizing (24) around $(\bar{Q},\bar{P},\bar{D})$ gives

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

where q_t, p_{t+1}, and d_{t+1} are deviations of $\log Q_t$, $\log P_{t+1}$, and $\log D_{t+1}$ from their means. The coefficient matrices are given by:

$$\begin{aligned} A_0^C &= \mathrm{Diag}(\bar{Q})^{-1}\frac{R^f}{\gamma}\Sigma^{-1}\mathrm{Diag}(\bar{P}), \\ A_1^C &= \mathrm{Diag}(\bar{Q})^{-1}\frac{1}{\gamma}\Sigma^{-1}\mathrm{Diag}(\bar{P}), \\ B_1^C &= \mathrm{Diag}(\bar{Q})^{-1}\frac{1}{\gamma}\Sigma^{-1}\mathrm{Diag}(\bar{D}). \end{aligned} \tag{25}$$

Because $R^f = \beta^{-1}$, the current- and future-price coefficients satisfy

$$A_1^C = \beta A_0^C. \tag{26}$$

Thus the benchmark CARA model is a special case of the general notation in Section 2 with $S = 1$.

4.2 Equilibrium with persistent outside flows

To isolate the price channel, suppose expected dividends are constant so that the term $B_1^C \mathbb{E}_t[d_{t+1}]$ (expected log deviations from mean) equals zero. Using (16), equilibrium is

$$A_0^C p_t - A_1^C \mathbb{E}_t[p_{t+1}] = \Lambda z_t. \tag{27}$$

Assume outside holdings follow

$$z_{t+1} = \rho z_t + \varepsilon_{t+1} + \eta_t, \qquad \rho \in [0,1), \tag{28}$$

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

Proposition 2 (Unique equilibrium)**.** *Suppose prices satisfy the no-bubble restriction*

$$\lim_{j\to\infty} \beta^j \mathbb{E}_t[p_{t+j}] = 0$$

for every t. Then the equilibrium condition (27) *admits the unique solution*

$$p_t = \mathcal{M} z_t + \beta \mathcal{M} \eta_t, \qquad \mathcal{M} = \left(A_0^C - \rho A_1^C\right)^{-1} \Lambda. \tag{29}$$

Using (25) *and* (26)*, this can be written as*

$$\mathcal{M} = \frac{1}{1-\beta\rho}(A_0^C)^{-1}\Lambda = \frac{\gamma\beta}{1-\beta\rho} Diag(\bar{P})^{-1}\Sigma. \tag{30}$$

Proposition 2 shows that equilibrium prices are linear in realized and expected outside holdings, and the coefficient for the latter is slightly lower by a fraction of β due to timing differences. The multiplier depends on the contemporaneous demand slope A_0^C, but also on the feedback from future expected prices induced by persistent outside flows.

The matrix $\mathcal{M}$ is the object we use throughout the paper. It is the log-price multiplier matrix: it maps percentage-of-shares demand shocks into percentage price responses.[4] If outside investors buy one percent of shares outstanding in asset j only, then equilibrium log prices rise by $0.01\,\mathcal{M}e_j$, that is, by the jth column of $0.01\,\mathcal{M}$. More generally, if outside investors buy one percent in some direction v, the response is $0.01\,\mathcal{M}v$, and if they buy one percent of shares outstanding in every asset, the response is $0.01\,\mathcal{M}\mathbf{1}$. This is why $\mathcal{M}$ is the natural object for counterfactuals stated in percentage-of-shares units.

4.3 Instrumented elasticity

In the general notation of Section 3, the benchmark is a stationary environment, so the multiplier matrix and the IE matrix are time-invariant. We therefore write them as $\mathcal{M}$ and $\mathcal{E}$ in this section.

Because (16) implies $dq_t/dz_t^\top = -\Lambda$, the IE with respect to the flow innovation is

$$\mathcal{E} = -\frac{dq_t}{dz_t^\top}\left(\frac{dp_t}{dz_t^\top}\right)^{-1} = \Lambda\mathcal{M}^{-1} = A_0^C - \rho A_1^C = (1-\beta\rho)A_0^C. \tag{31}$$

Define the corresponding aggregate elasticity as $\tilde{\mathcal{E}} \equiv \Lambda^{-1}\mathcal{E}$. Then

$$\tilde{\mathcal{E}} = \mathcal{M}^{-1}.$$

[4] Because total shares outstanding are normalized to 1, the level deviation z_t represents the absolute change in the fraction of market shares held by outsiders, making it the natural percentage-of-shares unit.

Hence the inverse multiplier is not the structural elasticity A_0^C. It is the *aggregate instrumented* elasticity, while the investor-level IE equals $\Lambda\mathcal{M}^{-1}$.

4.4 What does a flow instrument identify?

The benchmark equilibrium clarifies the connection to the taxonomy in Section 3. When outside holdings are persistent, a flow instrument cannot generate a purely static or purely immediate perturbation. Under (28), a shock to z_t changes not only current prices but also the entire expected path $\{\mathbb{E}_t[p_{t+s}]\}_{s\geq 1}$. In return language, it therefore moves the whole expected-return term structure rather than only next-period expected returns.

This is why the object in (31) 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 benchmark model that distinction is especially transparent: the structural slope is A_0^C, the investor-level IE equals $(1-\beta\rho)A_0^C$, and the aggregate dynamic IE equals the inverse multiplier.

4.5 Comparative statics

Equation (30) delivers the comparative statics used in the empirical section.

Proposition 3 (Own-price impact and return variance)**.** *Holding fixed* $(\gamma,\beta,\rho,\bar{P})$ *and all off-diagonal covariance terms, the own-price multiplier for asset* j *satisfies*

$$\mathcal{M}_{jj} = \frac{\gamma\beta}{1-\beta\rho}\frac{\Sigma_{jj}}{\bar{P}_j}.$$

Hence the own-price impact of a flow shock in asset j *is increasing in that asset's perceived return variance* Σ_{jj}.

Proposition 4 (Systematic versus idiosyncratic directions)**.** *Let v be a unit eigenvector of $Diag(\bar{P})^{-1}\Sigma$ with eigenvalue λ. A unit flow innovation in direction v generates price response*

$$\mathcal{M}v = \frac{\gamma\beta}{1-\beta\rho}\lambda v.$$

Hence percentage price impact is larger along directions associated with larger eigenvalues of $Diag(\bar{P})^{-1}\Sigma$.

Proposition 5 (Surprise versus expected flows)**.** *Suppose we decompose the one-period change in outside position into surprise and expected components:*

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

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:

$$B_{surp} = \mathcal{M}$$
$$B_{exp} = \mathcal{M} \cdot \underbrace{[W + (1-\beta)\Sigma_\eta] \cdot [W + \Sigma_\eta]^{-1}}_{\textit{attenuation factor}}$$

where $W \equiv \frac{1-\rho}{1+\rho}(\Sigma_\epsilon + \Sigma_\eta)$.

To gain intuition, consider the special case where $\Sigma_\epsilon = \sigma^2_\epsilon I_N$ and $\Sigma_\eta = \sigma^2_\eta I_N$. Then the attenuation factor becomes a scalar — $\frac{\frac{1-\rho}{1+\rho}(\sigma^2_\epsilon+\sigma^2_\eta)+(1-\beta)\sigma^2_\eta}{\frac{1-\rho}{1+\rho}(\sigma^2_\epsilon+\sigma^2_\eta)+\sigma^2_\eta}$ — 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.[5]

Proposition 6 (Flow persistence)**.** *Holding fixed* $(\gamma,\beta,\Sigma,\bar{P})$*, the multiplier matrix* $\mathcal{M}$ *is increasing in* ρ*.*

Discussions. These propositions have similar underlying intuition. A more volatile asset is harder for the arbitrageur to absorb, so a given percentage-point change in outside demand requires a larger percentage price response to clear the market. In the own-price dimension, this statement is exact: the multiplier is proportional to perceived return variance (Proposition 3). The same logic explains the systematic-direction result in Proposition 4. A flow aligned with a high-eigenvalue direction of $\mathrm{Diag}(\bar{P})^{-1}\Sigma$ loads on a combination of assets that is especially costly for the arbitrageur to warehouse. That direction therefore commands a larger multiplier. By contrast, an idiosyncratic direction with a low eigenvalue is easier to absorb and moves prices less.

In terms of dynamics, Proposition 5 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. In proposition 6, more persistent flows are also harder to absorb because today's imbalance predicts future imbalances; the arbitrageur therefore demands compensation not just for the current position but for the expected continuation of that position. This is exactly the $(1-\beta\rho)^{-1}$ amplification in (30). The higher is ρ, the more strongly a current flow shock propagates into future expected prices and returns, and the larger is the current price response.

Taken together, these comparative statics are exactly what make instrumented elasticities context-dependent. The multiplier matrix is not universal; it depends on risk, persistence, and

[5]To put numbers into perspective, at an annual frequency, we expect $\rho \approx \beta \approx 0.96$, in which case the attenuation factor becomes approximately $\frac{\sigma_\epsilon^2+2\sigma_\eta^2}{\sigma_\epsilon^2+51\sigma_\eta^2}$.

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.

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.[6] 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

[6] 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).

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 (Lou, 2012; Ben-David, Li, Rossi, and Song, 2022, e.g.) 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:

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

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.[7] 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

[7]OFI only captures all 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.

captures a subset of trades by mutual funds and ETFs, OFI captures all aggressively executed trading in the U.S. stock market and thus have 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 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 3). 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 3.

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. Column (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; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5.3 Systematic flows have larger multipliers than idiosyncratic flows

To test the prediction in Proposition 4 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

downloaded from Chen and Zimmermann (2021) and Fama-French 12 industry dummy indicators.[8] 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 4 implies that $M_{\text{systematic}} > M_{\text{idiosyncratic}}$.

The regressions 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.[9]

[8] 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.

[9] 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; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5.4 More surprising flows have larger multipliers

We now test Proposition 5 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 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:

$$\begin{aligned} f_{i,t}^{\text{surp}} &= f_{i,t} - \hat{f}_{i,t}^{\text{granular}} \\ f_{i,t}^{\text{less surp}} &= \hat{f}_{i,t}^{\text{granular}} - \hat{f}_{i,t}^{\text{coarse}} \\ f_{i,t}^{\text{predicted}} &= \hat{f}_{i,t}^{\text{coarse}} \end{aligned} \tag{32}$$

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.[10]

To approximate the belief of investors who learn in real time, in each year y, we use data in years $y-4,...,y-1$ to build a model and compute predictions on year y.[11] 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 (32), to test whether more surprising demand has larger price impact, we estimate multipliers via regression

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

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

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

where τ_t denote period-specific effects and we control for 13 weekly lags of past returns to control 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 $f_{i,t}^{\text{less surp}}$ $f_{i,t}^{\text{predicted}}$ 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 6 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 ($\varepsilon_{t,t+k}^{\text{exp}}$) is extracted. We then estimate a regression to see if price multipliers increases 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 is 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 ρ_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 denote the detrended component of $Z_{i,t}$ as $Z^{\text{det}}_{i,t}$. 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 ρ_i, written as $(\beta_0 + \beta_1 \rho_i)$, where β_0 and β_1 are freely estimated parameters. We estimate the regression:

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

where α_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^{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 ρ_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 6. 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. All specifications absorb time FEs and cluster by time. Standard errors are reported in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Results. Table 6 reports the results of estimating equation (5.5) 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 ρ_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 general equilibrium condition (17), restated here for convenience:

$$A_0 p_t - \sum_{s=1}^{S} A_s \mathbb{E}_t[p_{t+s}] = y_t, \qquad y_t \equiv \Lambda z_t + f_t. \tag{33}$$

Suppose the composite shifter y_t has the Wold representation

$$y_t = \bar{y} + \sum_{k=0}^{\infty} \Psi_{y,k} \varepsilon_{t-k}, \qquad \Psi_{y,0} = I_N,$$

and equilibrium prices admit the representation

$$p_t = \bar{p} + \sum_{k=0}^{\infty} \Xi_k \varepsilon_{t-k}. \tag{34}$$

Here Ξ_k denotes the reduced-form price impulse response at horizon k. Taking expectations in (34),

$$\mathbb{E}_t[p_{t+s}] = \bar{p} + \sum_{k=0}^{\infty} \Xi_{k+s}\varepsilon_{t-k}.$$

Substituting into (33) and matching coefficients on ε_{t-k} yields

$$A_0\Xi_k - \sum_{s=1}^{S} A_s\Xi_{k+s} = \Psi_{y,k}, \qquad k \geq 0. \tag{35}$$

We refer to (35) as the *recovery equation*: it must hold at every horizon k and therefore links the structural demand matrices $\{A_s\}$ to the observable impulse responses $\{\Xi_k, \Psi_{y,k}\}$.

6.2 Recovery under finite-support flow shocks

To isolate the role of demand shocks, suppose that the non-demand shifter f_t is constant, so that $\Psi_{y,k} = \Lambda\Psi_{z,k}$, where $\Psi_{z,k}$ denotes the impulse response of outside holdings to an identified demand innovation. Suppose further that the demand shock has finite support $L \geq S$, where S is the maximum horizon over which structural demand looks forward (i.e. $s = 1, \ldots, S$ in the demand system), meaning

$$\Psi_{z,k} = 0 \qquad \text{for all } k > L,$$

and that the induced price impulse responses also vanish beyond L, i.e. $\Xi_k = 0$ for all $k > L$. Economically, finite support means that the demand disturbance is transitory: it perturbs holdings for at most L periods before fully dissipating. This is natural for many empirical demand shocks—for example, a mutual fund rebalancing in response to investor flows generates trading concentrated in a finite window. The requirement $L \geq S$ ensures that the shock persists long enough to produce

impulse responses at every horizon over which structural demand is forward-looking, which is what makes the recursive identification below possible.

At horizon $k = L$, the recovery equation (35) reduces to

$$A_0 \Xi_L = \Lambda \Psi_{z,L},$$

because all terms involving horizons beyond L vanish. Hence

$$A_0 = \Lambda \Psi_{z,L} \Xi_L^{-1}. \tag{36}$$

At horizon $k = L - 1$, the recovery equation gives

$$A_0 \Xi_{L-1} - A_1 \Xi_L = \Lambda \Psi_{z,L-1},$$

so that

$$A_1 = \left(A_0 \Xi_{L-1} - \Lambda \Psi_{z,L-1} \right) \Xi_L^{-1}.$$

Proceeding recursively, for $j = 1, \ldots, S$,

$$A_j = \left(A_0 \Xi_{L-j} - \Lambda \Psi_{z,L-j} - \sum_{\ell=1}^{j-1} A_\ell \Xi_{L-j+\ell} \right) \Xi_L^{-1}. \tag{37}$$

Proposition 7 (Recovery from impulse responses)**.** *Suppose $L \geq S$, Ξ_L is invertible, and the demand shock is identified so that $\Psi_{z,0}, \ldots, \Psi_{z,L}$ and $\Xi_0, \ldots, \Xi_L$ are observed. Then the structural demand matrices $(A_0, \ldots, A_S)$ are identified recursively by* (36) *and* (37)*.*

Proof. Equation (36) identifies A_0. Given $(A_0, \ldots, A_{j-1})$, equation (37) identifies A_j because Ξ_L is invertible. Repeating the recursion identifies all matrices through A_S. □

6.3 Simple parameterizations for empirical implementation

The three dynamic specifications below form a nested sequence. The one-step model restricts demand to depend only on one-period-ahead expected returns. The two-step model allows sensitivity to expected returns at two horizons, with the restriction that the second-horizon coefficient is proportional to the first. The geometric specification extends this proportionality to an arbitrary horizon while keeping the number of free parameters small enough for weekly estimation. Throughout this subsection, the subscripts jj and $j,-j$ refer to the own and average-other entries of the symmetric 2×2 coefficient matrix; the empirical regressions that follow revert to stock index i and horizon index h.

The recursive result above is fully general but requires estimating separate matrices at each horizon. In practice, we impose low-dimensional restrictions on the *return-based* coefficients C_s (in (7)) rather than estimating the price coefficients A_s and B_s (in (8) freely, because the C_s are the primitive demand objects and the price and dividend coefficients are determined by them through the Campbell-Shiller restrictions (9). Throughout, we work in the symmetric own-versus-average-other basis identified by the weekly cross-sectional design, so each horizon-s coefficient is summarized by one diagonal term and one average off-diagonal term:

$$C_s \equiv \begin{pmatrix} C_{s,jj} & C_{s,j,-j} \\ C_{s,j,-j} & C_{s,jj} \end{pmatrix}, \qquad \tilde{C}_s \equiv \begin{pmatrix} \tilde{C}_{s,jj} & \tilde{C}_{s,j,-j} \\ \tilde{C}_{s,j,-j} & \tilde{C}_{s,jj} \end{pmatrix}.$$

Let Ψ_h^d and Ψ_h^o denote the own and average-other flow responses to a unit shock, let Ξ_h^d and Ξ_h^o denote the corresponding cumulative return responses, and let Ω_h^d and Ω_h^o denote the own and average-other idiosyncratic-volatility responses when continuation risk is included.

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 continuation-risk terms, then market clearing implies

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

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 $L = 0$ special case of the recovery equation, because then $A_0 \Xi_0 = I$ under the normalized shock. In the own/other basis, the naive benchmark therefore estimates only two parameters,

$$A_0 \equiv \begin{pmatrix} A_{0,jj} & A_{0,j,-j} \\ A_{0,j,-j} & A_{0,jj} \end{pmatrix},$$

which, after normalization into the aggregate elasticity units used in the empirical section, reduce to the two elements of $\mathcal{M}^{-1}$.

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

$$\Psi_h^d = C_{1,jj}\left(\Xi_h^d - \beta \Xi_{h+1}^d\right) + C_{1,j,-j}\left(\Xi_h^o - \beta \Xi_{h+1}^o\right),$$
$$\Psi_h^o = C_{1,j,-j}\left(\Xi_h^d - \beta \Xi_{h+1}^d\right) + C_{1,jj}\left(\Xi_h^o - \beta \Xi_{h+1}^o\right),$$

for $h = 0, \ldots, H$. This restriction has exactly two free return-demand parameters, $(C_{1,jj}, C_{1,j,-j})$. If continuation risk is added in the same one-step form, the system becomes

$$\Psi_h^d = C_{1,jj}\left(\Xi_h^d - \beta\Xi_{h+1}^d\right) + C_{1,j,-j}\left(\Xi_h^o - \beta\Xi_{h+1}^o\right) - \tilde{C}_{1,jj}\Omega_{h+1}^d - \tilde{C}_{1,j,-j}\Omega_{h+1}^o,$$
$$\Psi_h^o = C_{1,j,-j}\left(\Xi_h^d - \beta\Xi_{h+1}^d\right) + C_{1,jj}\left(\Xi_h^o - \beta\Xi_{h+1}^o\right) - \tilde{C}_{1,j,-j}\Omega_{h+1}^d - \tilde{C}_{1,jj}\Omega_{h+1}^o,$$

so the variance-augmented one-step model has four free parameters: $(C_{1,jj}, C_{1,j,-j}, \tilde{C}_{1,jj}, \tilde{C}_{1,j,-j})$.

Example 3: two-step recovery. Suppose $S = 2$, so that investors look two periods ahead. A fully unrestricted two-step system would allow separate matrices C_1 and C_2,

$$C_1 \equiv \begin{pmatrix} C_{1,jj} & C_{1,j,-j} \\ C_{1,j,-j} & C_{1,jj} \end{pmatrix}, \qquad C_2 \equiv \begin{pmatrix} C_{2,jj} & C_{2,j,-j} \\ C_{2,j,-j} & C_{2,jj} \end{pmatrix}.$$

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:

$$(C_{1,jj}, C_{1,j,-j}, \delta).$$

The implied flow equations are

$$\Psi_h^d = C_{1,jj}\left(\Xi_h^d - \beta\Xi_{h+1}^d\right) + C_{1,j,-j}\left(\Xi_h^o - \beta\Xi_{h+1}^o\right) + \delta C_{1,jj}\left(\Xi_{h+1}^d - \beta\Xi_{h+2}^d\right) + \delta C_{1,j,-j}\left(\Xi_{h+1}^o - \beta\Xi_{h+2}^o\right),$$

$$\Psi_h^o = C_{1,j,-j}\left(\Xi_h^d - \beta\Xi_{h+1}^d\right) + C_{1,jj}\left(\Xi_h^o - \beta\Xi_{h+1}^o\right) + \delta C_{1,j,-j}\left(\Xi_{h+1}^d - \beta\Xi_{h+2}^d\right) + \delta C_{1,jj}\left(\Xi_{h+1}^o - \beta\Xi_{h+2}^o\right).$$

If continuation risk is added with the same two-step shape, $\tilde{C}_2 = \delta\tilde{C}_1$, then the model has five free parameters:

$$(C_{1,jj}, C_{1,j,-j}, \delta, \tilde{C}_{1,jj}, \tilde{C}_{1,j,-j}).$$

Example 4: geometric continuation in the C coefficients. Extending the two-step restriction to an arbitrary 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 continuation risk is included, the same low-dimensional restriction on the risk coefficients,

$$\tilde{C}_s = \delta^{s-1}\tilde{C}_1, \qquad s = 1, \ldots, S.$$

In the same own/other basis, this implies

$$\Psi_h^d = \sum_{s=1}^{S} \delta^{s-1}\left[C_{1,jj}\left(\Xi_{h+s-1}^d - \beta\Xi_{h+s}^d\right) + C_{1,j,-j}\left(\Xi_{h+s-1}^o - \beta\Xi_{h+s}^o\right) - \tilde{C}_{1,jj}\Omega_{h+s}^d - \tilde{C}_{1,j,-j}\Omega_{h+s}^o\right],$$

$$\Psi_h^o = \sum_{s=1}^{S} \delta^{s-1}\left[C_{1,j,-j}\left(\Xi_{h+s-1}^d - \beta\Xi_{h+s}^d\right) + C_{1,jj}\left(\Xi_{h+s-1}^o - \beta\Xi_{h+s}^o\right) - \tilde{C}_{1,j,-j}\Omega_{h+s}^d - \tilde{C}_{1,jj}\Omega_{h+s}^o\right].$$

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

$$(C_{1,jj}, C_{1,j,-j}, \delta).$$

With continuation risk, it has five:

$$(C_{1,jj}, C_{1,j,-j}, \delta, \tilde{C}_{1,jj}, \tilde{C}_{1,j,-j}).$$

These four examples correspond exactly to the rows of Table 8 in the empirical result in Section 6.4: naive static $\mathcal{M}^{-1}$, one-step C, two-step C, and geometric C, each with and without the continuation-risk augmentation. For the two-step and geometric specifications, the shape parameter δ is estimated by one-dimensional minimum distance: conditional on a candidate δ, the own and average-other coefficients are obtained by least squares, and δ is then chosen to minimize the distance between model-implied and estimated impulse responses.[12]

6.4 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, so that both FIT and OFI discipline the same latent demand system on the same stock-weeks; the longer unbalanced OFI sample serves as a robustness check. Because both measures are already scaled by lagged shares outstanding, they are in the same units as the outside-demand shock z_t in the model.

The identified weekly demand shock is the *surprise* component of flow. In the baseline

[12] Appendix A writes down the estimation objective and the corresponding sequence of estimation steps used for the weekly recovery tables.

specification, for each measure we estimate a pooled stock-demeaned AR(4),

$$z_{i,t} = \sum_{\ell=1}^{4} \phi_\ell z_{i,t-\ell} \; + \; \sum_{\ell=1}^{4} \kappa_\ell r_{i,t-\ell} \; + \; \varepsilon_{i,t}, \tag{38}$$

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

$$z_{i,t+h} = \alpha^d_{t,h} + \Psi^d_h \varepsilon_{i,t} + \Gamma_h^\top X_{i,t} + u^d_{i,t+h}, \tag{39}$$

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

$$\sum_{j=0}^{h} r_{i,t+j} = a^d_{t,h} + \Xi^d_h \varepsilon_{i,t} + \Theta_h^\top X_{i,t} + v^d_{i,t+h}, \tag{40}$$

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

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 Xi,t contains foud lags of flows and returns:

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

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.

Measure	$\mathcal{M}_{jj}$	$\mathcal{M}_{j,-j}$	$\mathcal{M}_{jj}^{-1}$
FIT	4.278	-0.0012	0.234
	(0.258)	(0.0001)	(0.014)
OFI	2.544	-0.0006	0.393
	(0.083)	(0.0000)	(0.013)
Pooled overlap	2.587	-0.0007	0.387
	(0.040)	(0.0000)	(0.006)

Table 7. Weekly Multiplier Matrix in Own/Other Form.
The table reports the contemporaneous own and average-other price responses to the weekly surprise shock on the common 2007–2018 FIT–OFI overlap sample. For each measure, the shock is the residual from the stock-demeaned AR(4) in (38). The pooled overlap row aggregates the same own/other multiplier object and therefore matches the naive benchmark used in Table 8. Standard errors are Newey-West across weekly Fama-MacBeth coefficient series for the FIT and OFI rows and are delta-method transforms of the overlap benchmark reported in Table 8.

This own-versus-average-other split is motivated by the identification logic in Haddad, He, Huebner, Kondor, and Loualiche (2025a). Cross-sectional designs are informative about relative own-versus-cross effects, but time-series variation is needed to discipline substitution. We therefore estimate a symmetric low-dimensional system directly in the own/other basis. For the contemporaneous multiplier, the maintained restriction is

$$\mathcal{M} \approx \begin{pmatrix} \mathcal{M}_{jj} & \mathcal{M}_{j,-j} \\ \mathcal{M}_{j,-j} & \mathcal{M}_{jj} \end{pmatrix},$$

where the off-diagonal element is the average spillover to other stocks from a unit shock to stock j.

Table 7 reports these reduced-form objects on the common 2007–2018 FIT–OFI overlap sample. The pooled overlap row is the apples-to-apples benchmark for Table 8: its diagonal multiplier is about 2.59, so the corresponding naive inverse benchmark is about 0.39. We call these "naive" because they shut down all continuation terms and equate the structural elasticity with the contemporaneous inverse multiplier alone.

To recover structural demand, we treat FIT and OFI as *proxy measures* for a common latent

demand disturbance and estimate a single low-dimensional set of return-demand coefficients from their joint own/other impulse responses. Pooling is important here, not merely convenient. On overlapping stock-weeks the two surprise shocks are only weakly aligned: their correlation is 0.046, FIT explains only about 0.2% of OFI surprise variation, and the two measures agree in sign only about 52% of the time. FIT is therefore a narrow and noisy proxy for the underlying disturbance, whereas OFI is broader but less transparently interpretable.

For this reason, we do not interpret the FIT-only and OFI-only recoveries as competing structural estimates. Because each proxy loads on the common disturbance with a different signal-to-noise ratio, separate recoveries can differ simply due to proxy-specific attenuation rather than instability in the underlying demand system. The pooled estimator fits a single set of coefficients to both sets of impulse responses jointly, weighting each proxy by its relative precision. Disagreement across the shock-specific estimates is therefore informative about proxy quality, not about the structural parameters themselves.

Table 8 reports the pooled recovery results on the common 2007–2018 FIT–OFI sample.[13] The first row is the naive static benchmark based on M^{-1} alone. In own/other form it delivers a diagonal elasticity of only about 0.39, because it effectively assumes away all future expected-return and continuation-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 4.25, and adding continuation-risk moments increases it further to about 6.16. 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 for the more persistent parameterizations. The two-step specification gives

[13] Appendix Table A.1 repeats the same recovery exercise on the longer unbalanced samples as a robustness check.

Spec	$C_{1,jj}$	$C_{1,j,-j}$	Shape	$\tilde{C}_{1,jj}$	$\tilde{C}_{1,j,-j}$
Naive static $\mathcal{M}^{-1}$	0.387 (0.006)	0.0001 (< 0.0001)	–	–	–
One-step C only	4.254 (0.798)	> −0.0001 (< 0.0001)	–	–	–
One-step C + var.	6.162 (1.300)	< 0.0001 (0.0001)	–	11.128 (12.336)	0.0023 (0.0008)
Two-step C only	5.318 (1.028)	-0.0001 (0.0001)	0.333 (0.277)	–	–
Two-step C + var.	5.837 (1.263)	> −0.0001 (0.0001)	0.485 (0.282)	13.295 (10.319)	0.0018 (0.0008)
Geometric C only	5.360 (1.182)	-0.0001 (0.0001)	0.229 (0.329)	–	–
Geometric C + var.	5.599 (1.365)	> −0.0001 (0.0001)	0.492 (0.320)	13.446 (8.722)	0.0017 (0.0008)

Table 8. Structural Recovery from Weekly Own/Other Impulse Responses.
The table reports pooled structural coefficients recovered from the weekly own and average-other impulse responses on the common 2007–2018 FIT–OFI 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 continuation-risk moments based on idiosyncratic-volatility impulse responses. The "Shape" column reports the decay parameter, δ, for the two-step and geometric specifications. Standard errors in parentheses are obtained by resampling weeks. Shock-specific FIT and OFI recoveries are used only as diagnostics, because the two proxy shocks are only weakly correlated on overlapping stock-weeks. Full-sample robustness results are reported in Appendix Table A.1.

$\hat{C}_{1,jj} \approx 5.32$ without continuation risk and $\hat{C}_{1,jj} \approx 5.84$ with it. The geometric specification gives $\hat{C}_{1,jj} \approx 5.36$ and $\hat{C}_{1,jj} \approx 5.60$, respectively. Across all six dynamic specifications, the average off-diagonal elasticity 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. Davis and Chaudhry (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 same direction: the dominant structural margin is own-stock demand, and the average

Measure	$\Omega_{jj,1}$	$\Omega_{j,-j,1}$
FIT	0.0043	-0.00000
	(0.0051)	(0.00000)
OFI	0.0013	-0.00000
	(0.0004)	(0.00000)

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

offsetting reallocation into other stocks is close to zero. In our specification, separate parameters govern own-price and cross-price elasticities within each period's coefficient matrix, which is more flexible than logit-style demand, where a single parameter controls both. A single cross-elasticity parameter 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 volatility moments that discipline the continuation-risk augmentation of Table 8. The key fact is that the own one-week-ahead idiosyncratic-volatility response is positive for both FIT and OFI, and statistically significant for OFI, whereas the average-other response is close to zero. This pattern is exactly the one needed for the variance-augmented models to push the recovered diagonal elasticity upward: if the same shock that raises prices also worsens near-term risk-bearing conditions in the shocked stock, then the observed price response must be reconciled with a larger underlying expected-return elasticity. This covariance/risk effect has been emphasized by He et al. (2025), and empirically we find this 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 continuation-risk 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 five or more, with little evidence of average cross-stock substitution in the own/other basis.

These results clarify what the data can and cannot reveal. A single stock-level shock is not enough to identify a fully unrestricted elasticity matrix with rich substitution. But the weekly data do pin down a disciplined low-dimensional object: a common diagonal structural elasticity that is much larger than $\mathcal{M}^{-1}$, an off-diagonal elasticity that is close to zero in average-other form, and an economically meaningful increase in the diagonal coefficient once continuation risk is allowed to move with the shock.

6.5 Interpretation

Proposition 7 shows why the log-linear rewrite is useful. The same coefficients that describe structural demand in Section 2 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. Persistent instruments are therefore not a nuisance. They are exactly what makes 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 under a weak price-mediated exclusion restriction 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 0.3 to 0.4, whereas the recovered structural elasticity is about 4.3 to 6.2. 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.

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.

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.

Cochrane, John H, 2008, The dog that did not bark: A defense of return predictability, *The Review of Financial Studies* 21, 1533–1575.

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

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

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

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.

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.

Li, Yizhang, Stanislav Sokolinski, and Andrea Tamoni, 2026, Which investors drive anomaly returns and how?, *Journal of Financial Economics* 179, 104257.

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

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 Additional Weekly Recovery Results

For the weekly recovery tables, estimation proceeds in two steps. First, we estimate the empirical own and average-other impulse responses, collected in a vector $\hat{m}$, from the local projections in (39) and (40), and when continuation risk is included we append the corresponding volatility moments. Second, for a given parameter vector θ, we compute the model-implied moments $m(\theta)$ under the relevant restriction from Section 6. The reported estimates solve the minimum-distance problem

$$\hat{\theta} \in \arg\min_{\theta} \ \big(\hat{m} - m(\theta)\big)^{\top} W \big(\hat{m} - m(\theta)\big),$$

where W is a diagonal weighting matrix based on the relative sample precision of the FIT and OFI moment blocks. In the naive static and one-step specifications, the model is linear in θ, so the estimator reduces to least squares. In the two-step and geometric specifications, δ is the only nonlinear parameter: conditional on a candidate δ, the remaining own and average-other coefficients are estimated by least squares, and then δ is chosen by a one-dimensional search over the minimum-distance objective. Standard errors in the main text and appendix are obtained by resampling weeks and re-estimating the full procedure.

B Proofs for Instrumented Elasticity Results

Proof of Proposition 1. By linearized market clearing (16),

$$q_t = -\Lambda z_t.$$

Differentiating with respect to z_t gives

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

Spec	$C_{1,jj}$	$C_{1,j,-j}$	Shape	$\tilde{C}_{1,jj}$	$\tilde{C}_{1,j,-j}$
Naive static $\mathcal{M}^{-1}$	0.346 (0.003)	0.0001 (< 0.0001)	–	–	–
One-step C only	4.414 (0.742)	< 0.0001 (0.0001)	–	–	–
One-step C + var.	5.482 (0.999)	< 0.0001 (0.0001)	–	2.195 (35.669)	0.0041 (0.0036)
Two-step C only	4.996 (0.821)	-0.0001 (0.0001)	0.384 (0.264)	–	–
Two-step C + var.	5.152 (0.932)	> −0.0001 (0.0001)	0.447 (0.275)	11.495 (31.897)	0.0033 (0.0032)
Geometric C only	5.037 (0.863)	-0.0001 (0.0001)	0.274 (0.300)	–	–
Geometric C + var.	4.995 (0.922)	< 0.0001 (0.0001)	0.630 (0.343)	22.661 (31.279)	0.0037 (0.0030)

Table A.1. Full-Sample Robustness for Weekly Structural Recovery.
The table repeats Table 8 on the longer unbalanced weekly samples rather than the common 2007–2018 overlap sample. This appendix table is intended as a robustness check because FIT and OFI do not span the same stock-weeks outside the overlap period. Standard errors in parentheses are obtained by resampling weeks.

By definition, the price multiplier matrix is

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

Assumption (i) implies that $\mathcal{M}_t(z_t)$ is invertible, so substituting these two expressions into (13) yields

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

This proves (20).

Now define the aggregate demand elasticity matrix

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

Using (20),

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

which proves (21). Price-mediated exogeneity is the additional interpretation condition ensuring that the matrix in (13) is the IE of interest rather than a mixture of price-mediated and direct characteristic effects. □

C Proofs for Benchmark CARA Results

Proof of Proposition 2. Because (26) implies

$$A_0^C - \rho A_1^C = (1 - \beta\rho)A_0^C,$$

and A_0^C is invertible while $1 - \beta\rho > 0$, the matrix $A_0^C - \rho A_1^C$ is invertible. Define

$$\mathcal{M} \equiv \left(A_0^C - \rho A_1^C\right)^{-1} \Lambda.$$

Since $\mathbb{E}_t[z_{t+1}] = \rho z_t$, the process $p_t = \mathcal{M}z_t$ satisfies

$$A_0^C p_t - A_1^C \mathbb{E}_t[p_{t+1}] = \left(A_0^C \mathcal{M} - \rho A_1^C \mathcal{M}\right) z_t = \Lambda z_t,$$

so it is an equilibrium.

To prove uniqueness, let $\tilde{p}_t$ be any other equilibrium satisfying the same no-bubble restriction and define $u_t \equiv \tilde{p}_t - \mathcal{M}z_t$. Subtracting the equilibrium condition for $\mathcal{M}z_t$ from that for $\tilde{p}_t$ yields

$$A_0^C u_t - A_1^C \mathbb{E}_t[u_{t+1}] = 0.$$

Using (26), this becomes

$$u_t = \beta\mathbb{E}_t[u_{t+1}].$$

Iterating forward gives

$$u_t = \beta^j \mathbb{E}_t[u_{t+j}] \qquad \text{for all } j \geq 1.$$

Because both $\tilde{p}_t$ and $\mathcal{M}z_t$ satisfy the no-bubble restriction,

$$\lim_{j\to\infty} \beta^j \mathbb{E}_t[u_{t+j}] = 0.$$

Taking limits in the iterated equation therefore gives $u_t = 0$ almost surely for all t, so $\tilde{p}_t = \mathcal{M}z_t$. □

Proof of Proposition 3. From (30),

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

Taking the (j,j) entry gives

$$\mathcal{M}_{jj} = \frac{\gamma\beta}{1-\beta\rho}\frac{\Sigma_{jj}}{\bar{P}_j}.$$

Holding fixed $(\gamma,\beta,\rho,\bar{P})$ and all off-diagonal covariance terms, the own-price multiplier therefore increases one-for-one with Σ_{jj}. □

Proof of Proposition 6. Equation (30) scales with $(1-\beta\rho)^{-1}$, which is increasing in ρ on $[0,1)$. □

Proof of Proposition 4. If v is a unit eigenvector of $\text{Diag}(\bar{P})^{-1}\Sigma$ with eigenvalue λ, then by (30),

$$\mathcal{M}v = \frac{\gamma\beta}{1-\beta\rho}\text{Diag}(\bar{P})^{-1}\Sigma v = \frac{\gamma\beta}{1-\beta\rho}\lambda v.$$

Hence the magnitude of the directional price response is increasing in λ. □

Proof of Proposition 5. From (29), $p_t = \mathcal{M}z_t = \mathcal{M}(\rho z_{t-1} + \varepsilon_t)$, so

$$\mathbb{E}_{t-1}[p_t] = \mathcal{M}\rho z_{t-1} \qquad \text{and} \qquad p_t - \mathbb{E}_{t-1}[p_t] = \mathcal{M}\varepsilon_t.$$

□

D Equivalent representations

This appendix uses a more generic functional notation to show how to move between alternative representations of the same demand system. The purpose is purely algebraic: it records the

equivalence mappings behind the main text. We start with the representation in terms of price and dividends:

$$q_t = q^{(p,d)}(p_t, \mathbb{E}_t[p_{t+1}], \mathbb{E}_t[d_{t+1}], \mathbb{E}_t[p_{t+2}], \mathbb{E}_t[d_{t+2}], \dots; f_t).$$

From the Campbell-Shiller approximation, we obtain

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

To simplify the notation, we set $\kappa = 0$. Results extend trivially to the case where $\kappa \neq 0$. We will also omit the demand shifter f_t for brevity.

Equivalence between (p,d) and (p,r). Solving for expected dividends at horizon h we obtain:

$$\mathbb{E}_t[d_{t+h}] = \frac{\mathbb{E}_t[r_{t+h}] + \mathbb{E}_t[p_{t+h-1}] - \rho \mathbb{E}_t[p_{t+h}]}{1-\rho}.$$

We can use the expression above to eliminate expected dividends from the demand system:

$$\begin{aligned} q_t &= q^{(p,r)}(p_t, \mathbb{E}_t[p_{t+1}], \mathbb{E}_t[r_{t+1}], \mathbb{E}_t[p_{t+2}], \mathbb{E}_t[r_{t+2}], \dots) \\ &= q^{(p,d)}(p_t, \mathbb{E}_t[p_{t+1}], \frac{\mathbb{E}_t[r_{t+1}] + p_t - \rho \mathbb{E}_t[p_{t+1}]}{1-\rho}, \mathbb{E}_t[p_{t+2}], \frac{\mathbb{E}_t[r_{t+2}] + \mathbb{E}_t[p_{t+1}] - \rho \mathbb{E}_t[p_{t+2}]}{1-\rho}, \dots), \end{aligned}$$

where we omitted the constant κ for brevity. This implies that the structural demand elasticities of one representation are connected to the structural demand elasticities of another representation:

$$\frac{\partial q_t^{(p,r)}}{\partial \mathbb{E}_t[r_{t+h}]} = \frac{1}{1-\rho} \frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[d_{t+h}]}, \qquad \frac{\partial q_t^{(p,r)}}{\partial \mathbb{E}_t[p_{t+h}]} = \frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[p_{t+h}]} - \frac{\rho}{1-\rho} \frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[d_{t+h}]} + \frac{1}{1-\rho} \frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[d_{t+h+1}]},$$

for $h \geq 1$, and

$$\frac{\partial q_t^{(p,r)}}{\partial p_t} = \frac{\partial q_t^{(p,d)}}{\partial p_t} + \frac{1}{1-\rho} \frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[d_{t+1}]}.$$

We can perform the reverse operation, that is, recover the (p,d) representation from the (p,r) representation. We start again from the Campbell-Shiller approximation:

$$\mathbb{E}_t[r_{t+h}] = \rho \mathbb{E}_t[p_{t+h}] + (1-\rho) \mathbb{E}_t[d_{t+h}] - \mathbb{E}_t[p_{t+h-1}].$$

We can then use the expression above to eliminate expected returns from the demand system in (p,r):

$$\begin{aligned} q_t &= q^{(p,d)}(p_t, \mathbb{E}_t[p_{t+1}], \mathbb{E}_t[d_{t+1}], \mathbb{E}_t[p_{t+2}], \mathbb{E}_t[d_{t+2}], \ldots) \\ &= q^{(p,r)}(p_t, \mathbb{E}_t[p_{t+1}], \rho\mathbb{E}_t[p_{t+1}] + (1-\rho)\mathbb{E}_t[d_{t+1}] - p_t, \mathbb{E}_t[p_{t+2}], \rho\mathbb{E}_t[p_{t+2}] + (1-\rho)\mathbb{E}_t[d_{t+2}] - \mathbb{E}_t[p_{t+1}], \ldots). \end{aligned}$$

The structural demand elasticities of the (p,d) representation are connected to the structural demand elasticities of the (p,r) representation:

$$\frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[d_{t+h}]} = (1-\rho)\frac{\partial q_t^{(p,r)}}{\partial \mathbb{E}_t[r_{t+h}]}, \qquad \frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[p_{t+h}]} = \frac{\partial q_t^{(p,r)}}{\partial \mathbb{E}_t[p_{t+h}]} + \rho\frac{\partial q_t^{(p,r)}}{\partial \mathbb{E}_t[r_{t+h}]} - \frac{\partial q_t^{(p,r)}}{\partial \mathbb{E}_t[r_{t+h+1}]},$$

for $h \geq 1$, and

$$\frac{\partial q_t^{(p,d)}}{\partial p_t} = \frac{\partial q_t^{(p,r)}}{\partial p_t} - \frac{\partial q_t^{(p,r)}}{\partial \mathbb{E}_t[r_{t+1}]}.$$

Equivalence between (p,d) and (r,d). We now show that the (p,d) representation is equivalent to the (r,d) representation. We start again from the Campbell-Shiller approximation, but this time we solve the equation forward and use a no-bubble condition:

$$p_t = \sum_{j=1}^{\infty} \rho^{j-1} \left[(1-\rho)\mathbb{E}_t[d_{t+j}] - \mathbb{E}_t[r_{t+j}]\right].$$

Evaluating at horizon h and taking expectations, we obtain:

$$\mathbb{E}_t[p_{t+h}] = \sum_{j=h+1}^{\infty} \rho^{j-(h+1)} \left[(1-\rho)\mathbb{E}_t[d_{t+j}] - \mathbb{E}_t[r_{t+j}]\right] \equiv g_{t,h}(\mathbb{E}_t[r_{t+h+1}], \mathbb{E}_t[d_{t+h+1}], \ldots).$$

We can then use the expression above to eliminate expected returns from the demand system in (p,d):

$$\begin{aligned} q_t &= q^{(r,d)}(\mathbb{E}_t[r_{t+1}], \mathbb{E}_t[d_{t+1}], \mathbb{E}_t[r_{t+2}], \mathbb{E}_t[d_{t+2}], \ldots) \\ &= q^{(p,d)}(\underbrace{g_{t,0}(\mathbb{E}_t[r_{t+1}], \mathbb{E}_t[d_{t+1}], \ldots)}_{p_t}, \underbrace{g_{t,1}(\mathbb{E}_t[r_{t+2}], \mathbb{E}_t[d_{t+2}], \ldots)}_{\mathbb{E}_t[p_{t+1}]}, \mathbb{E}_t[d_{t+1}], \ldots). \end{aligned}$$

We can then use the expression above to recover the structural demand elasticities of the (r,d) representation from the structural demand elasticities of the (p,d) representation:

$$\frac{\partial q_t^{(r,d)}}{\partial \mathbb{E}_t[r_{t+h}]} = \sum_{j=0}^{h-1} \frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[p_{t+j}]} \frac{\partial g_{t,j}}{\partial \mathbb{E}_t[r_{t+h}]}, \qquad \frac{\partial q_t^{(r,d)}}{\partial \mathbb{E}_t[d_{t+h}]} = \sum_{j=0}^{h-1} \frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[p_{t+j}]} \frac{\partial g_{t,j}}{\partial \mathbb{E}_t[d_{t+h}]} + \frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[d_{t+h}]}.$$

for $h \geq 1$.

We can perform the reverse operation, that is, recover the (p,d) representation from the (r,d) representation. We start again from the Campbell-Shiller approximation:

$$\mathbb{E}_t[r_{t+h}] = \rho \mathbb{E}_t[p_{t+h}] + (1-\rho)\mathbb{E}_t[d_{t+h}] - \mathbb{E}_t[p_{t+h-1}].$$

We can then use the expression above to eliminate expected returns from the demand system in (r,d):

$$\begin{aligned} q_t &= q^{(p,d)}(p_t, \mathbb{E}_t[p_{t+1}], \mathbb{E}_t[d_{t+1}], \mathbb{E}_t[p_{t+2}], \mathbb{E}_t[d_{t+2}], \ldots) \\ &= q^{(r,d)}(\rho \mathbb{E}_t[p_{t+1}] + (1-\rho)\mathbb{E}_t[d_{t+1}] - p_t, \mathbb{E}_t[d_{t+1}], \rho \mathbb{E}_t[p_{t+2}] + (1-\rho)\mathbb{E}_t[d_{t+2}] - \mathbb{E}_t[p_{t+1}], \mathbb{E}_t[d_{t+2}], \ldots). \end{aligned}$$

We can then use the expression above to recover the structural demand elasticities of the (p,d) representation from the structural demand elasticities of the (r,d) representation:

$$\frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[p_{t+h}]} = \rho \frac{\partial q_t^{(r,d)}}{\partial \mathbb{E}_t[r_{t+h}]} - \frac{\partial q_t^{(r,d)}}{\partial \mathbb{E}_t[r_{t+h+1}]}, \qquad \frac{\partial q_t^{(p,d)}}{\partial \mathbb{E}_t[d_{t+h}]} = (1-\rho)\frac{\partial q_t^{(r,d)}}{\partial \mathbb{E}_t[r_{t+h}]} + \frac{\partial q_t^{(r,d)}}{\partial \mathbb{E}_t[d_{t+h}]}.$$

for $h \geq 1$, and

$$\frac{\partial q_t^{(p,d)}}{\partial p_t} = -\frac{\partial q_t^{(r,d)}}{\partial \mathbb{E}_t[r_{t+1}]}.$$

Conclusion. We have seen that the (p,d) representation is equivalent to the (p,r) representation and the (r,d) representation. Hence, the (p,r) and (r,d) representations are also equivalent. This implies that our results on recovering the structural demand elasticities from the data are valid for all three representations.