Endogenous Elasticities: Price Multipliers Are Smaller for Larger Demand Shocks *

Aditya Chaudhry and Jiacui Li

February 12, 2026

ABSTRACT

How well can markets absorb large demand shocks? We address this question using price multipliers: the per-unit price adjustment investors require to absorb shocks. We find stock-level multipliers are smaller for larger contemporaneous and cumulative past shocks. Using holdings data, we find investors endogenously become more price elastic as larger shocks create larger price dislocations. These results are consistent with models where investors become more willing to absorb shocks when large dislocations create large profit opportunities, such as fixed adjustment costs or endogenous inattention. Our findings suggest large shocks expand rather than exhaust the capacity of markets to absorb shocks.

JEL classification: G11, G12
Keywords: Demand-System Asset Pricing, Price Multipliers

*Chaudhry (chaudhry.127@osu.edu) is at the Fisher College of Business at The Ohio State University. Li (jiacui.li@eccles.utah.edu) is at the David Eccles School of Business, University of Utah. We are grateful for comments from Hengjie Ai, Yu An, Ehsan Azarmsa, Itzhak Ben-David, Carter Davis, Espen Eckbo, Julie Fu, Xaviar Gabaix, Michael Gallmeyer, Niels Gormsen, Valentin Haddad, Lars Hansen, Da Huang, Paul Huebner, Paymon Khorrami, Ralph Koijen, Erik Loualiche, Federico Mainardi, Stefan Nagel, Sangmin Oh, Anna Pavlova, Taisiya Sikorskaya, Philippe Van Der Beck, Kaushik Vasudevan, Song Xiao, as well as participants at Princeton University, NBER Asset Pricing Fall 2025, UNC Junior Conference, Hong Kong Poly U, Hong Kong University, MARC, NYU WAPFIN, Stanford SITE, The Ohio State University, Utah, and UVA McIntire.

1 Introduction

How well can financial markets absorb large demand shocks to individual assets? The asset pricing literature offers two views. In one view, large shocks exhaust absorption capacity. For example, in models with leverage and short-sale constraints, liquidity provider risk aversion, or limited arbitrage capital, large shocks tighten constraints and render investors less able to absorb additional quantities (Amihud and Mendelson, 1980; Kyle and Xiong, 2001; Gromb and Vayanos, 2002; Hong and Stein, 2003; Brunnermeier and Pedersen, 2009). Under this view, markets easily absorb small shocks, but large shocks create large price dislocations. In the other view, however, large shocks *expand* absorption capacity. For example, in models with fixed adjustment costs or endogenous inattention, large shocks create large profit opportunities that incentivize investors to overcome their frictions and absorb these shocks (Lo et al., 2004; Liu, 2004; Van Nieuwerburgh and Veldkamp, 2010; Gabaix, 2019). Under this view, large shocks attract stabilizing capital that dampens the growth of price dislocations with shock size. Understanding which view dominates is essential to understanding the price impact of the large demand shocks that are often of interest, such as government asset purchases (e.g., buying 10% of Intel), green investing, or share buybacks and issuances.

In this paper, we show the latter view dominates in the cross section of stocks: large shocks expand absorption capacity, which dampens the growth of price dislocations with shock size.

To assess which view of absorption capacity holds empirically, we measure price multipliers: the per-unit price impact of demand shocks (Gabaix and Koijen, 2022). Price multipliers quantify how much compensation investors require to absorb these shocks. While previous work finds that price multipliers are large for relatively small demand shocks, it is unclear how multipliers scale to the large shocks that are often of interest. We document a new stylized fact: price multipliers are smaller for larger contemporaneous and cumulative past shocks. These results suggest that investors endogenously become more price elastic — more willing to absorb shocks — as larger shocks create larger price dislocations and profit

opportunities. We confirm this interpretation in holdings data by showing that institutional investors become more price-elastic following larger price dislocations. Overall, our findings support the view that absorption capacity expands in response to large shocks.

We first provide empirical evidence that larger uninformed demand shocks have smaller price multipliers (in Section 4). To estimate multipliers, we use three demand shocks from the literature: Russell index reconstitution-induced changes in benchmarking intensity (BMI) from Pavlova and Sikorskaya (2023), fund flow-induced trading (FIT) from Lou (2012), and order flow imbalance (OFI) from Li and Lin (2025). We acknowledge that no measure is perfect because identifying purely exogenous variation in demand is difficult, especially given our focus on large shocks. Nevertheless, previous work has studied these shocks extensively and has shown them to be largely unrelated to cash flow news.

Using these three shocks, we find that price multipliers are smaller for larger demand shocks. We divide shocks into bins by shock magnitude and run cross-sectional price impact regressions at monthly and quarterly frequencies. For instance, when studying the price impact of BMI, we find that demand shocks below one standard deviation in magnitude (i.e., those smaller than 0.73% of shares outstanding) have multipliers around 2. That is, buying X% of shares outstanding raises prices by around 2X%. Yet for shocks above two standard deviations in magnitude (i.e., those above 1.46% of shares outstanding), multipliers are less than 1, which is statistically and economically significantly smaller. We also find similar patterns for the other two demand shocks. In addition to these results based on *contemporaneous* shock size, we also find that multipliers decrease with the magnitude of cumulative shocks over the past several quarters (in Section 4.3).

A potential concern is that the multiplier variation we document arises from other differences between large and small shocks, or from differences between stocks experiencing large versus small shocks, beyond just shock size itself. We address several variants of this concern. First, we estimate all regressions cross-sectionally within each time period, mitigating concerns that

large demand shocks cluster in periods when price multipliers are lower. Second, we control for cross-sectional differences in liquidity, addressing concerns that large shocks concentrate in more liquid stocks, which may have smaller multipliers. Third, we show that differences across shock magnitudes in anticipation, in correlation with omitted demand shocks, and in return reversal speed do not explain the variation in multipliers we document.

We next examine institutional holdings data to shed light on the origins of this price multiplier variation (in Section 5). We estimate price elasticities of demand — which are inversely related to multipliers — for institutional investors in the U.S. stock market using 13F holdings data. Motivated by our preceding findings, we allow elasticities to depend on the magnitude of price dislocations. Using the optimal granular instrumental variable (GIV) approach of Chaudhary et al. (2024), which extends the approach of Gabaix and Koijen (2024), we find that most institutions indeed become more price-elastic following larger price dislocations. This behavior is particularly pronounced for more active institutions.

Lastly, in Section 6 we discuss which mechanisms can rationalize our empirical results. Our findings are consistent with mechanisms in which investors face frictions that make their demand inelastic, but can overcome these frictions to endogenously become more elastic when price dislocations are large. We consider two examples of this type of mechanism. First, this behavior arises in models with fixed adjustment costs (e.g., Grossman and Laroque (1987); Lo et al. (2004)), which give rise to (S, s)-style adjustment behavior (Scarf, 1960). For small shocks, an investor's initial position remains close to optimal, so the fixed cost exceeds the benefit of adjustment. Large shocks, however, significantly alter the optimal position by altering expected returns, making adjustment worthwhile and investors more elastic. Second, this behavior also arises in models with endogenous inattention (e.g., Sims (2003); Van Nieuwerburgh and Veldkamp (2009, 2010); Kacperczyk et al. (2016); Gabaix (2019)) where uncertainty about cash flows or expected returns renders demand inelastic. If investors can pay to acquire information to reduce this uncertainty, they will be more

willing to do so as demand shocks accumulate and create larger profit opportunities. Lower uncertainty renders investors' demand more elastic.

Our findings have important implications for evaluating how shifts in quantities impact financial markets. Price multipliers for the large demand shocks relevant to many finance questions — such as government asset purchases, green investing, and share buybacks and issuances — are likely smaller than multiplier estimates based on small shocks. Moreover, our findings provide guidance for future work that seeks to model investors' asset demand. While previous work on asset demand systems often assumes fixed price elasticities, our results suggest that incorporating the endogenous behavior of these elasticities can allow for a more realistic characterization of demand and price dynamics. Doing so can thus provide more accurate answers to questions about how shifts in quantities affect asset prices.

We clarify several points of interpretation regarding our results. First, although price multipliers decrease with shock size, they remain substantially larger than predicted by frictionless models.[1] Thus, our findings shed light on endogenous variation in price multipliers rather than suggest that multipliers are as small as in frictionless asset pricing models. Second, our findings are distinct from the "square root" price impact curves at the trade level documented in the microstructure literature (e.g., Frazzini et al., 2018). Those studies focus on small trades at high-frequency, intraday time horizons. In contrast, our analysis examines much larger shocks over longer horizons of interest for asset pricing (several months or quarters), at which most microstructure effects have dissipated. Third, since our empirical setting is the cross section of stocks, there is an open question of how multipliers vary with shock size in other markets and for other levels of aggregation.

Apart from our empirical findings, we note that economic intuition also suggests that price

[1]Frictionless asset pricing models predict that uninformed demand shocks have very low stock-level price multipliers. Specifically, Petajisto (2009) predict multipliers on the order of 1/6,000. Davis et al. (2025) shows that even if one accounts for the fact that the passthrough of price movements to expected returns is less than one, the predicted multiplier increases only to around 1/800.

multipliers may decline with shock size. Empirical work estimates stock-level multipliers to be around 2.[2] These estimates appear reasonable for small shocks; for example, uninformed selling of 1% of shares outstanding lowers prices by $1 - e^{2\times(-1\%)} \approx 2\%$. However, if we consider uninformed selling of 30% of shares (over an extended period of time), assuming a constant multiplier implies that price will decline by $1 - e^{2\times(-30\%)} \approx 45\%$, which seems large if there is no change in fundamentals. It seems plausible that some investors may increase their demand for the stock before such large undervaluation can materialize. Therefore, it seems likely that price multipliers decline for shocks that are sufficiently large.

The paper is structured as follows. Section 2 introduces our empirical test for determining how market absorption capacity varies with shock size: testing how price multipliers vary with shock size. Section 3 describes our data and demand shock measures. Section 4 presents empirical evidence that price multipliers decrease with the magnitude of both current and cumulative past demand shocks. Section 5 analyzes institutional holdings data and shows that price elasticities increase with the magnitude of price dislocations. Section 6 discusses the mechanisms that can generate our empirical results. Section 7 concludes.

1.1 Related Literature

This paper contributes to the literature on market absorption capacity and investor frictions. One strand of this literature emphasizes frictions that lead large shocks to exhaust absorption capacity. For example, models emphasizing leverage constraints (Gromb and Vayanos, 2002; Brunnermeier and Pedersen, 2009) and short-selling constraints (Diamond and Verrecchia, 1987; Duffie et al., 2002; Hong and Stein, 2003) predict that large shocks tighten investors' constraints and reduce their capacity to absorb quantities, creating large price dislocations. In contrast, a different strand of the literature emphasizes frictions that relax as price dislocations grow, and so lead large shocks to *expand* absorption capacity. For example, models with fixed adjustment costs (Constantinides, 1986; Grossman and Laroque, 1987) or endoge-

[2] See Gabaix and Koijen (2022) for a review of the literature.

nous inattention (Grossman and Stiglitz, 1980; Kyle, 1989; Sims, 2003; Van Nieuwerburgh and Veldkamp, 2009, 2010; Kacperczyk et al., 2016; Gabaix, 2019) predict that large profit opportunities attract stabilizing capital. We show that the latter class of frictions dominates in the cross section of stocks: large shocks expand absorption capacity, which dampens the growth of price dislocations with shock size.

Moreover, this paper contributes to the growing demand system asset pricing literature that estimates the effects of uninformed demand shocks on asset prices. This literature, using both plausibly exogenous demand shocks (Shleifer (1986); Harris and Gurel (1986); Lou (2012); Chang et al. (2015); Hartzmark and Solomon (2022); Gabaix and Koijen (2022); Schmickler and Tremacoldi-Rossi (2022); Pavlova and Sikorskaya (2023); An et al. (2025); Jansen (2025)) and structural asset demand systems (Koijen and Yogo (2019, 2020); Han et al. (2021); Huebner (2023); Haddad et al. (2024); Jansen et al. (2024); Koijen et al. (2024); Tamoni et al. (2024); Bretscher et al. (2025)), finds that price elasticities of demand are empirically much smaller and price multipliers are much higher than predicted by frictionless asset pricing models. Given the difficulty in finding natural experiments involving large demand shocks, this literature has focused on plausibly exogenous but relatively small shocks. We contribute to this literature by shedding light on how price multipliers scale to the large shocks that are often of interest. We find multipliers endogenously decline with the magnitude of current and past demand shocks, and establish mechanisms that give rise to this pattern.

While the literature has generally not focused on endogenous variation in multipliers, a notable exception is Haddad et al. (2024), who find that elasticities vary across *stocks* with different investor bases due to strategic investor interactions. In contrast, we show that elasticities vary across *shocks* with different magnitudes, representing a different source of multiplier variation. Additionally, He et al. (2025) show theoretically that demand shocks can alter elasticities in general equilibrium. The dominant channel in their calibration — endogenous risk — predicts multipliers increase with shock size, opposite to our findings.

Their model also features an intertemporal hedging channel that is directionally consistent with our results, but cannot quantitatively rationalize our findings in their calibration.

Additionally, the market microstructure literature documents that price impacts follow a "square-root law" at high frequencies, such as at the trade or order level, consistent with our finding that multipliers decrease with shock size (e.g., Tóth et al., 2011). Correspondingly, the explanations offered focus on microstructure. In the economics literature, researchers explain this pattern with mechanisms based on liquidity supplier risk or cost (e.g., first-order risk aversion in the face of inventory uncertainty, as in Gabaix et al., 2006) and the time required to offset trades (e.g., a monopolistic market maker competing with stochastic order flow, as in Chacko et al., 2008). Explanations have also been offered in the econophysics literature (Alfonsi et al., 2010; Gatheral, 2010). For instance, Donier et al. (2015) derives square-root price impact under the assumption of local linearity in the density of limit orders. Bouchaud et al. (2018) argues that price impact tends to be concave in frequently traded markets, but closer to linear in infrequently traded markets. Our work is distinct as we focus on lower frequencies: monthly and quarterly. To our knowledge, no previous work has established whether the nonlinearity documented in microstructure settings extends to the lower frequencies and the much larger demand shocks we analyze. Crucially, because we study frequencies at which microstructure effects have likely dissipated, the potential explanations for our findings likely differ from those in the microstructure literature.

2 Assessing Absorption Capacity with Multipliers

To study how well financial markets can absorb large demand shocks, we assess how price multipliers endogenously vary with shock size.

Price multipliers are the per-unit price impact of demand shocks (Gabaix and Koijen, 2022). That is, if some investors buy or sell $\Delta Q\%$ of the shares outstanding of a given stock for

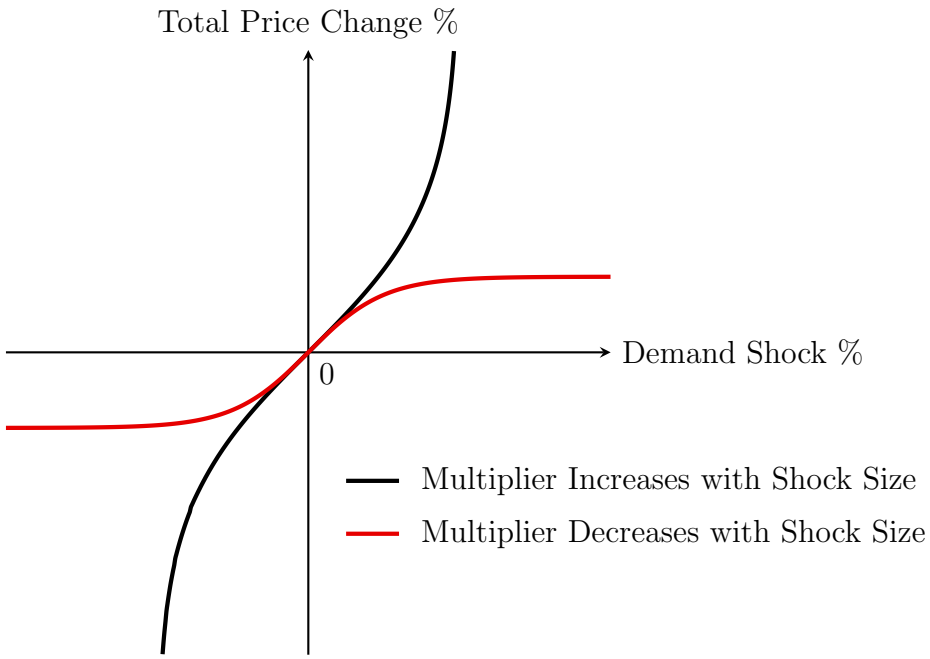


Figure 1. Convex (black) and concave (red) total price impact curves corresponding to price multipliers increasing versus decreasing with shock size, respectively.

reasons unrelated to cash flow news, and price changes by $\Delta P\%$, then the price multiplier is

$$M \equiv \frac{\%\Delta P}{\%\Delta Q}.$$

The multiplier reflects how much prices must adjust to incentivize other investors to absorb this demand shock. Large multipliers suggest that investors require significant compensation to absorb additional quantities, and so indicate limited absorption capacity.

Hence, how price multipliers vary with shock size reveals whether large shocks exhaust or expand absorption capacity.

If large shocks exhaust absorption capacity, then price multipliers should *increase* with shock size. This behavior arises in models with leverage constraints (e.g., Gromb and Vayanos (2002); Brunnermeier and Pedersen (2009)) and short sale constraints (e.g., Hong and Stein (2003)). Intuitively, larger shocks tighten investors constraints and render them less able to absorb additional quantities. Although price dislocations and expected returns are large

and investors want to take larger positions, they cannot due to constraints. Thus, as shocks grow, investors become less price-elastic and multipliers rise. In these models, total price impact is convex in shock size, as the black curve in Figure 1 displays .

By contrast, if large shocks expand absorption capacity, then multipliers should *decrease* with shock size. This behavior arises in models with fixed adjustment costs (e.g., Lo et al. (2004)) and endogenous inattention (e.g., Van Nieuwerburgh and Veldkamp (2010); Gabaix (2019)). Intuitively, frictions render investors unable or unwilling to absorb shocks. However, large shocks create large price dislocations and large profit opportunities that incentivize investors to overcome these frictions. Thus, as shocks grow, investors endogenously become more price-elastic — more willing to absorb shocks — and multipliers decrease. In these models, total price impact is concave in shock size, as the red curve in Figure 1 displays.

3 Data and Demand Shocks

To estimate price multipliers, we use plausibly exogenous asset demand shocks that are not correlated with cash flow news.[3] Specifically, we use three demand shocks from previous work that researchers argue are largely devoid of cash flow information. We recognize the difficulty in finding cash flow-unrelated demand shocks and that each shock has its own strengths and limitations. For this reason, we analyze three different shocks and highlight that our results prove robust across all three.

3.1 Data

We download monthly stock returns and market capitalizations from CRSP, and aggregate monthly returns to the quarterly frequency when needed. To control for known return predictors, all regressions control for 13 commonly used stock characteristics from Chen and Zimmermann (2022): accruals, asset growth, beta, book-to-market, gross profitability,

[3]Otherwise, the price change associated with a quantity change would reflect the price impacts of both the quantity change and the correlated cash flow news, leading to omitted variable bias in multiplier estimates.

industry momentum, intermediate momentum, 1 year issuance, 5 year issuance, momentum, seasonal momentum, net operational assets, and short-term reversal. We also control for six liquidity proxies which include size, effective bid-ask spread, quoted bid-ask spread, realized volatility, turnover, and dollar trading volume. These variables are obtained from CRSP and WRDS Intraday Indicators. Following common practice in the literature on modeling the cross-section of returns, we transform each characteristic into a uniform distribution over [-0.5, 0.5] in each cross-section (e.g. Kelly et al., 2019).

To construct the demand shocks, we obtain data for benchmarking intensity (BMI) from Pavlova and Sikorskaya (2023). For flow-induced-trading (FIT), we use mutual fund flows from the CRSP Mutual Fund database, holdings from the Thomson Reuters S12 database, and link the two using MFLINKS from Russ Wermers. For order flow-imbalance (OFI), we use daily Lee-Ready classified order flows from WRDS intraday indicators.

The elasticity estimation in Section 5 uses institutional holdings data from the SEC Form 13F (from 1980 to 2021) provided by Thomson Reuters through WRDS. The SEC requires all institutional investors with at least $100 million in assets under management (AUM) to report stock-level long positions each quarter. Following Koijen and Yogo (2019), we construct a residual "household" sector that comprises direct holdings by households and those by non-13F institutions (i.e., institutions with less than $100 million AUM).

3.2 Demand Shocks

We briefly explain the construction and logic behind the three demand shocks we use. For more details on these shocks and further arguments for why they are unrelated to cash flow news, we refer the reader to the original papers.

Benchmarking Intensity (BMI). Our first demand shock is based on index reconstitution, which captures changes in benchmarked investors' demand for a stock when it enters or exits an index (Shleifer (1986); Harris and Gurel (1986); Chang et al. (2015)). In particular,

we use changes in the benchmarking intensity (BMI) measure from Pavlova and Sikorskaya (2023), which provides a continuous measure of demand changes driven by Russell index reconstitutions. While there are also shocks based on other indices, such as the S&P 500 (e.g., Shleifer, 1986), Russell indices stand out for their mechanical inclusion criteria and BMI captures cross-sectional variation in shock magnitude (as a fraction of shares outstanding).

Russell index reconstitution occurs each June and is a mechanical function of market capitalization with no discretionary judgment. Each May, Russell ranks eligible stocks by market capitalization. Stocks above a specified rank cutoff join the Russell 1000, while those below join the Russell 2000.[4] The Russell 2000 has historically attracted more benchmarked institutional capital than the Russell 1000. Thus, a stock that starts in the Russell 1000 but falls below the cutoff in May moves to Russell 2000 and experiences inflows of benchmarked capital in June. Conversely, stocks that move from the Russell 2000 to 1000 experience outflows. Conditional on the market cap as of the May ranking date, Russell index membership in June is exogenous to June cash-flow news, and so these reconstitution-driven flows are uninformed demand shocks (Chang et al. (2015); Crane et al. (2016); Glossner (2019)).

Each reconstitution episode provides cross-sectional variation in demand shock magnitudes because stocks move across multiple Russell indices simultaneously. Every stock in the Russell 2000 Blend index is also in the Russell 2000 Value or Growth indices, which have different levels of benchmarked capital. Every stock in the Russell 1000 Blend index is also in the Russell 1000 Value or Growth indices, and some — those with market cap rank below 200 — are in the Russell Midcap Blend, Value, and Growth indices. Thus, a stock moving from the Russell 1000 Value to the 2000 Value experiences different inflows of benchmarked capital than a stock moving from the 1000 Growth to the 2000 Growth. This heterogeneity

[4]More specifically, prior to 2007 the rank cutoff was the 1,000th stock. To reduce turnover, since 2007 Russell has used a "banding policy" under which there are two separate cutoffs for stocks starting in the Russell 1000 and 2000 pre-reconstitution, both of which are still mechanical functions of the firm size distribution. Thus, there is a "band" of market caps including stocks from the Russell 1000 and 2000. Appendix A.1 explains the Russell methodology we use to calculate these cutoffs.

allows estimation of the relationship between price impact and shock size in each time period.

The Pavlova and Sikorskaya (2023) benchmarking intensity (BMI) measure captures this cross-sectional variation in Russell index reconstitution-induced demand shocks:

$$BMI_{n,t} = \sum_{\text{Index } j} \frac{\begin{array}{c}\text{Institutional AUM Benchmarked to Index } j \text{ in Month } t \\ \times \text{ Weight of Stock } n \text{ in Index } j \text{ in Month } t\end{array}}{\text{Stock } n \text{ Market Value in Month } t}.$$

BMI depends on which indices stock n is part of and on the proportion of n's market value that is held by benchmarked investors. BMI is calculated from thirty-four indices, including the nine Russell indices, that cover 90% of mutual fund and ETF assets under management.

We focus on exogenous changes in BMI due to the Russell index reconstitution each June. Specifically, we consider June BMI changes, denoted ΔBMI, for stocks in a narrow window around Russell 1000/2000 reconstitution thresholds (150 stocks on both sides of the threshold). Stocks with positive (negative) ΔBMI experience benchmarking inflows (outflows). While the *level* of BMI is generally correlated with stock profitability because index membership is based on market capitalization, the *change* ΔBMI for stocks in this window is driven by Russell index membership changes, which are exogenous to June cash flow news conditional on the May rank-date market cap.

We use the BMI and Russell index constituent data provided by Pavlova and Sikorskaya (2023). All of our specifications include the stock-level controls used by Pavlova and Sikorskaya (2023): May rank-date log market cap, one-year monthly average bid-ask percentage spread, and the banding controls from Appel et al. (2019).[5] Whereas Pavlova and Sikorskaya (2023) use the proprietary Russell market cap, we calculate market cap from standard databases using the method of Ben-David et al. (2019).[6]

[5]The banding controls include an indicator for rank-date market cap in the "band" with stocks from the Russell 1000 and 2000, an indicator for being in the 2000 in May, and the interaction of these indicators.

[6]Since Russell ranks stocks using a proprietary market cap that we lack access to, we use the method of Ben-David et al. (2019) to approximate this market cap using standard databases. Doing so accurately predicts assignment to the Russell indices. We use May, not June, market caps to calculate the reconstitution thresholds to avoid selection bias(e.g., Chang et al., 2015; Appel et al., 2021; Wei and Young, 2021).

Flow-Induced Trading (FIT). Our second demand shock is flow-induced trading (FIT) from Lou (2012), which captures the predicted, non-discretionary, stock-level trading of mutual funds and ETFs in response to flows.[7]

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 (Figure A4 in Li (2022)). Lou (2012) show that the price impact of FIT reverts over time, which suggests that such trades are not motivated by superior cash flow information.

While funds tend to scale existing positions proportionally in response to flows, the relationship between flows and trading can exhibit heterogeneity as a function of flow and pre-existing position sizes. Therefore, we construct a refined version of the Lou (2012) FIT measure to account for this heterogeneity and avoid potential mechanical biases in our results (Appendix A.2). We first calculate the quarterly percentage flow to fund i as

$$f_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1} \cdot (1 + \text{Ret}_{i,t})}{\text{TNA}_{i,t-1}}.$$

where $TNA_{i,t}$ and $\text{Ret}_{i,t}$ are fund i's end-of-quarter-t total net assets and return from end of quarter $t-1$ to t, respectively. The predicted non-discretionary trading by fund i in stock n due to this flow is proportional to $\text{SharesHeld}_{i,n,t-1} \cdot f_{i,t}$. To allow for heterogeneous flow-to-trade relationships, we estimate a scaling factor $b(w_{i,n,t-1}, f_{i,t})$ that is allowed to flexibly vary with realized fund flow $f_{i,t}$ and pre-existing portfolio weight $w_{i,n,t-1}$. Empirical estimates show that this scaling factor is approximately equal to one and varies little across funds. Appendix A.2 provides further details for the estimation of this scaling factor. Aggregating

[7] While not initially intended as a demand instrument, it has been used as such by subsequent work (e.g., Li, 2022; Chaudhry, 2023; Van der Beck, 2025).

across all funds and dividing by shares outstanding yields:

$$\text{FIT}_{n,t} = \sum_{\text{fund } i} b(w_{i,n,t-1}, f_{i,t}) \cdot \underbrace{\frac{\text{SharesHeld}_{i,n,t-1}}{\text{Shares Outstanding}_{n,t-1}}}_{\equiv S_{i,n,t-1}} f_{i,t}, \tag{1}$$

where $S_{i,n,t-1}$ is the proportion of stock n's shares owned by mutual fund i in quarter $t-1$.

Why is FIT unrelated to cash flow news? Lou (2012), and subsequently Li (2022) and Huang et al. (2024), argue that a large fraction of flows are directed by unsophisticated retail investors who do not respond to cash flow news. They also show that the price impact of FIT reverts over time, which suggests these shocks are unrelated to cash flow news. Chaudhry (2023) further argues that, because FIT is a shift-share instrument (Goldsmith-Pinkham et al., 2020), FIT can be unrelated to cash flow news even if flows are related to cash flow news. A weaker requirement is that the lagged ownership shares ($S_{i,n,t-1}$) of a given fund i are cross-sectionally uncorrelated with stock-level cash flow news after cross-sectionally controlling for various stock characteristics (as done in Section 4.1).

Order flow imbalance (OFI). Our third demand shock is 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 quarterly frequencies, and then normalize by lagged shares outstanding. 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. The OFI measure has been used heavily in the microstructure literature to study price impact at daily or higher frequencies. Recently, Li and Lin (2025) argue that OFI can also be useful for studying price impact at "asset pricing frequencies" (i.e., monthly or lower).

The benefit of OFI is that it has far more variation than the previous two shocks, as shown

in the summary statistics in Table 1. This large variation is particularly helpful for this paper because we are interested in price impact of large demand shocks.

At the same time, a drawback of OFI is its potential correlation with cash flow news. While BMI and FIT isolate trades by specific investors in specific circumstances that are unrelated to cash flow news, OFI captures trading by many investors. Li and Lin (2025) 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.

Comparing the three shocks. Each of our three demand shocks has strengths and weaknesses. BMI is arguably the most exogenous, but since index changes happen relatively rarely, it has the smallest sample size. Some researchers may be somewhat concerned about the exogeneity of FIT, but it provides more statistical power since it impacts all stocks in all periods. OFI likely has the greatest concerns regarding exogeneity, but it also provides the most power since it features the largest shocks. Given these tradeoffs, we use all three shocks and find our results are robust across all three.

				Percentiles						
	Obs	Mean	StDev	1%	5%	25%	50%	75%	95%	99%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Quarterly return (%)	5,000.6	2.85	27.32	-54.99	-35.18	-11.63	0.85	13.73	46.60	96.01
ΔBMI (%)	471.8	0.04	0.73	-1.97	-1.00	-0.23	0.01	0.27	1.26	2.18
FIT (%)	4,564.3	0.06	0.45	-0.96	-0.53	-0.15	0.03	0.24	0.76	1.42
OFI (%)	4,434.8	-0.58	4.02	-14.30	-6.32	-1.79	-0.31	0.99	4.47	9.42
Market cap ($m)	4,993.1	4,264	21,235	7	17	101	438	1,878	16,605	69,368

Table 1. Summary statistics for price multiplier estimation

This table reports time-series averages of cross-sectional moments. ΔBMI is changes in benchmarking intensity from Pavlova and Sikorskaya (2023), FIT is flow-induced trading from Lou (2012), and OFI is order flow imbalance from Li and Lin (2025). Column (1) reports the average number of stocks per period. For FIT and OFI, one period is one quarter. For ΔBMI, one period is the month of June — which is when Russell index reconstitution occurs — in a specific year. As described in the paper, we have rescaled BMI so it is expressed as a fraction of total shares outstanding. The sample for all variables consists of quarterly data from 1993 to 2022, except the ΔBMI sample consists of data for each June from 1998 through 2018. All demand shocks are expressed as a percentage of shares outstanding; for this reason, we multiply ΔBMI by 0.2, following Pavlova and Sikorskaya (2023).

Table 1 reports summary statistics. Since our empirical work exploits cross-sectional variation, the table reports time-series averages of cross-sectional moments. Because we study how price multipliers vary by demand shock magnitude, we show the percentile distributions of each shock in the right-most columns. All shocks are reported as a percentage of lagged shares outstanding. BMI and FIT have similar variation, with standard deviations of 0.73% and 0.45%, respectively.[8] OFI has the greatest variation with a standard deviation of 4.02%.

Our analysis spans 1998 to 2018 for the BMI demand shock, and 1993 to 2022 for FIT and OFI. The BMI sample is limited by the period for which we observe Russell index constituents. The OFI sample is limited by the availability of the signed OFI measure from WRDS. We constrain the FIT sample to match that of OFI.

4 Evidence from Price Multipliers

In this section, we demonstrate that price multipliers are smaller for larger shocks, which implies that large shocks expand market absorption capacity. Sections 4.1 and 4.2 present evidence that price multipliers are smaller for larger contemporaneous demand shocks. Section 4.3 shows parallel evidence based on the magnitude of cumulative past demand shocks.

4.1 Price Multipliers Increase with Contemporaneous Shock Size

We test how price multipliers vary with demand shock size. In each period, we sort demand shocks $d_{n,t}$ (ΔBMI, FIT, or OFI) into three bins based on size $|d_{n,t}|$: $b \in \{[0, \sigma_t), [\sigma_t, 2\sigma_t], [2\sigma_t, \infty)\}$, where σ_t is the cross-sectional standard deviation of $d_{n,t}$. We then estimate bin-specific multipliers $\{M_{b,t}\}_b$ using cross-sectional regressions of stock returns $r_{n,t}$ on demand shocks $d_{n,t}$

[8] Since ΔBMI measures the Russell index reconstitution-induced change in demand for a stock by only benchmarked mutual funds and ETFs — not the total change in demand by all institutional investors — we scale it to have the same units as OFI and FIT. Pavlova and Sikorskaya (2023) find that a 1% reconstitution-induced BMI increase raises total institutional ownership by 0.2%, so we multiply the raw ΔBMI by 0.2. See Column 3 of Table 3 in Pavlova and Sikorskaya (2023) for justification for this scaling.

interacted with bin indicators $I_{|d_{n,t}|\in b}$:

$$r_{n,t} = \sum_{\text{bin } b} M_{b,t} \cdot I_{|d_{n,t}|\in b} \cdot d_{n,t} + \boldsymbol{c}_t^{'} \boldsymbol{x}_{n,t-1} + \tau_t + \epsilon_{n,t} \tag{2}$$

The controls $\boldsymbol{x}_{n,t-1}$ include the stock characteristics discussed in Section 3.1, and the BMI-specific controls discussed in the same section when $d_{n,t}$ is ΔBMI. τ_t are time fixed effects.

We estimate these regressions at the quarterly frequency for FIT and OFI to capture low-frequency price impacts of interest for asset pricing. For BMI, to remain consistent with previous work (Pavlova and Sikorskaya, 2023), we use the monthly frequency and focus on price impact during the Russell reconstitution month of June.

We estimate regression (2) in each cross-section and report the time-series average regression coefficients, following the methodology of Fama and MacBeth (1973). This approach ensures we use only cross-sectional variation to identify the regression coefficients, thereby avoiding concerns that omitted variables drive time variation in both multipliers and shock sizes.

The $M_{b,t}$ coefficients should be interpreted as "relative" (Haddad et al., 2025) or "micro" (Gabaix and Koijen, 2022) multipliers.[9] We control for the set of return-predicting and liquidity stock characteristics discussed in Section 3.1 in each cross-section. Following Haddad et al. (2025), under the assumption of homogeneous substitution conditional on observables, these controls absorb substitution effects driven by characteristic-based comovement, allowing for the unbiased estimation of the relative multiplier.

Table 2 reports the regression results. For each of our three demand shocks, Panel A presents the bin-specific multiplier estimates and Panel B presents the multiplier differences between each pair of bins. Columns (1), (4), and (7) report the results without controls $\boldsymbol{x}_{n,t-1}$, while columns (2), (5), and (8) control for 13 return predictors. Columns (3), (6), and (9) further

[9]The relative multiplier is the difference between the own- and cross-price multipliers. It measures the price response of an asset relative to a one with the same characteristics due to a shift in relative demand.

Panel A: price impact regressions									
	Dependent variable: stock return $r_{n,t}$								
	BMI			FIT			OFI		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$M_{\lvert d_{n,t}\rvert<\sigma}$	2.44***	1.78***	1.77***	4.78***	3.96***	3.92***	2.82***	2.84***	3.13***
	(0.52)	(0.50)	(0.50)	(0.84)	(0.63)	(0.52)	(0.10)	(0.09)	(0.10)
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	1.41**	1.06*	1.05*	3.69***	3.17***	3.33***	1.86***	1.85***	2.03***
	(0.71)	(0.59)	(0.58)	(0.46)	(0.36)	(0.30)	(0.09)	(0.08)	(0.07)
$M_{\lvert d_{n,t}\rvert>2\sigma}$	0.88***	0.68**	0.69**	3.13***	2.67***	2.69***	0.81***	0.81***	0.91***
	(0.29)	(0.29)	(0.29)	(0.34)	(0.27)	(0.26)	(0.06)	(0.05)	(0.06)
Predictor controls	N	Y	Y	N	Y	Y	N	Y	Y
Liquidity controls	N	N	Y	N	N	Y	N	N	Y
Obs	9,910	9,910	9,910	561,404	561,404	561,404	527,744	527,744	527,744
R^2	0.051	0.142	0.172	0.009	0.064	0.080	0.053	0.106	0.129
Marginal $R^2(d_{n,t})$	0.019	0.014	0.014	0.009	0.006	0.005	0.053	0.047	0.049
Panel B: Coefficient differences									
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]} - M_{\lvert d_{n,t}\rvert<\sigma}$	−1.03	−0.72	−0.72	−1.10**	−0.78**	−0.59*	−0.96***	−0.99***	−1.10***
	(0.65)	(0.60)	(0.59)	(0.48)	(0.38)	(0.35)	(0.07)	(0.07)	(0.07)
$M_{\lvert d_{n,t}\rvert>2\sigma} - M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	−0.53	−0.38	−0.36	−0.56*	−0.50**	−0.64***	−1.05***	−1.04***	−1.12***
	(0.70)	(0.60)	(0.54)	(0.30)	(0.25)	(0.24)	(0.06)	(0.05)	(0.05)
$M_{\lvert d_{n,t}\rvert>2\sigma} - M_{\lvert d_{n,t}\rvert<\sigma}$	−1.56***	−1.10**	−1.08**	−1.65**	−1.29**	−1.23***	−2.01***	−2.03***	−2.22***
	(0.51)	(0.47)	(0.46)	(0.66)	(0.50)	(0.44)	(0.08)	(0.07)	(0.08)

Table 2. Price multipliers by contemporaneous demand sizes

Panel A reports results for Fama-MacBeth regressions:

$$r_{n,t} = \sum_{\text{bin } b} M_{b,t} \cdot I_{|d_{n,t}|\in b} \cdot d_{n,t} + \boldsymbol{c}_t' \boldsymbol{x}_{n,t-1} + \tau_t + \epsilon_{n,t} \tag{4}$$

where $r_{n,t}$ is the stock return and $d_{n,t}$ represents demand shock (ΔBMI, FIT, or OFI). In each period, we split the sample into three bins b by comparing $|d_{n,t}|$ against the cross-sectional standard deviation of $d_{n,t}$. Columns (1)-(3) report results using the ΔBMI demand shock and monthly returns. Columns (4)-(6) and (7)-(9) report results using quarterly returns and the FIT and OFI demand shocks, respectively. Columns (2), (5), and (8) include controls for 13 return-predicting stock characteristics. Columns (3), (6), and (9) further control for 6 liquidity-related proxies. These controls are described in Section 3.1. Panel B reports pairwise differences in price multipliers. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

control for the 6 liquidity proxies.

Across all specifications, the coefficients $\{M_b\}_b$ imply price multipliers are smaller for larger demand shocks. We focus on the specification with all controls in columns (3), (6), and (9). For BMI in column (3), the multiplier declines from 1.77 for shocks with magnitudes below one standard deviation, to 1.05 for magnitudes between one and two standard deviations, to 0.69 for magnitudes above two standard deviations. That is, price rises by $1.77X\%$ if a noise trader buys $X\%$ of shares outstanding for a stock for $0 < X < \sigma$, but only by $0.69X\%$ for $X > 2\sigma$. For FIT in column (6), the multiplier falls from 3.92 to 3.33 to 2.69 across bins.

For OFI in column (9), the multiplier falls from 3.13 to 2.03 to 0.91 across bins.[10]

This variation in price multipliers is economically significant; ignoring it overstates the price impact of large demand shocks. For example, consider the price impact of a two standard deviation OFI shock, which amounts to buying $2 \times 4.02\% = 8.04\%$ of shares outstanding of a stock. Our estimates in column (9) indicate the price impact would be $0.91\% \times 8.04 \approx 7.32\%$, whereas applying the small-shock multiplier would suggest a much larger price impact of $3.13\% \times 8.04 \approx 25.17\%$.

In addition to being economically significant, most of the multiplier differences across bins are statistically significant, as shown in Panel B of Table 2. For FIT and OFI, almost all pairwise multiplier comparisons are statistically significant at the 5%, if not 1%, level. Even though the BMI analysis has less statistical power as Russell index changes happen rarely (the sample size is less than 10,000 stock-months), the difference between the highest and lowest price multipliers is statistically significant at the 5% level.

Appendix Table B.3 confirms that estimating multipliers using panel regressions with time fixed effects yields qualitatively similar results.

Overall, these results suggest that as larger demand shocks create larger price dislocations, investors endogenously become more elastic, which expands market absorption capacity and lowers multipliers.

4.2 Addressing Alternative Interpretations

We interpret the results in Section 4.1 as evidence that, ceteris paribus, a stock exhibits a smaller price multiplier when experiencing a larger demand shock (a *within* stock comparison). However, the results may also be consistent with a general class of alternative interpretations: that price multipliers do not vary with shock size per se, but do vary for

[10] There are multiple potential reasons why the multipliers differ across demand shock types. For instance, ΔBMI primarily impacts stocks with smaller market capitalizations than FIT and OFI. Therefore, comparing multipliers across shock types is not a ceteris paribus exercise. Thus, we study variation within each type.

other reasons that correlate with shock size. In general, one may suspect that the true model exhibits linear price impact but with variation in price multipliers *across* stocks or time:

$$r_{n,t} = \underbrace{(b + \beta_{n,t})}_{\text{price multiplier}} \cdot d_{n,t} + \mathbf{c}_t' \boldsymbol{x}_{n,t-1} + \tau_t + \epsilon_{n,t}, \tag{5}$$

where $\beta_{n,t}$ denotes unobserved price multiplier variation that can correlate with $|d_{n,t}|$. This situation could create omitted variable bias that spuriously suggests multipliers decrease with shock size. We address several variants of this concern.

Time-varying multipliers. One potential concern is that multipliers happen to be larger in times when demand shocks are smaller. For instance, in the 2008 financial crisis, liquidity conditions worsened and price multipliers were larger ($\beta_{n,t}$ was larger), while at the same time there may also be less trading ($|d_{n,t}|$ was smaller). Our baseline specification already addresses this concern as we estimate (2) in each period using only cross-sectional variation.

Heterogeneous multipliers across stocks with different characteristics. Another potential concern is that multipliers are smaller for certain stocks — such as those that are more liquid — and these stocks also experience larger demand shocks. Formally, this concern can be modeled in (5) as $\beta_{n,t} = b_{3,t}' \cdot \boldsymbol{z}_{n,t-1}$ where $\boldsymbol{z}_{n,t-1}$ are stock-specific liquidity-related characteristics that may correlate with shock size $|d_{n,t}|$.

To address this concern, we augment our baseline specification (2) with interactions of the demand shock with stock characteristics $\boldsymbol{z}_{n,t-1}$:

$$r_{n,t} = \sum_{\text{bin } b} M_{b,t} \cdot I_{|d_{n,t}| \in b} \cdot d_{n,t} + \underbrace{\boldsymbol{b}_{3,t}' \cdot \boldsymbol{z}_{n,t-1} \cdot d_{n,t}}_{\text{Added}} + \boldsymbol{c}_t' \boldsymbol{x}_{n,t-1} + \tau_t + \epsilon_{n,t}. \tag{6}$$

where the interaction terms $\boldsymbol{b}_{3,t}' \cdot \boldsymbol{z}_{n,t-1} \cdot d_{n,t}$ absorb the $\beta_{n,t} \cdot d_{n,t}$ term in (5). As in Section

4.1, we estimate regression (6) in each cross-section and report the average coefficients, in the style of Fama and MacBeth (1973).

Table 3 reports the results. For comparison purposes, columns (1), (4), and (7) reproduce the specification with the full set of controls in Table 2. Columns (2), (5), and (8) also control for interactions of demand shocks with the 13 return predictors. Columns (3), (6), and (9) further control for interactions of demand shocks with 6 liquidity variables. The controls are described in Section 3.1. The results are qualitatively unchanged: we continue to find evidence that multipliers decrease with shock size.

Appendix Figure B.3 further shows that our results are robust throughout the process of progressively adding each individual characteristic interaction. Specifically, we plot the evolution of the pairwise multiplier differences (e.g., $M_{|d_{n,t}|\in[\sigma,2\sigma]} - M_{|d_{n,t}|<\sigma}$) as we add one interaction at a time, and find that the estimates are stable throughout all specifications. Therefore, these tests assuage the concern that our results are explained by cross-sectional correlations between stock characteristics and demand shock sizes.

Stock-specific variation in multipliers and demand shock sizes. To the extent that our stock characteristics capture the relevant variation of multipliers across stocks, Table 3 shows that differential characteristics across stocks do not explain our findings. However, one may worry that there is further unobserved variation of liquidity across stocks that we cannot control for. Specifically, let $\sigma_n(d_{n,t})$ be the stock-specific demand-shock volatility. Suppose cross-sectional variation in $\beta_{n,t}$ (multipliers) cannot be fully captured by stock characteristics. In this case, if stocks with lower $\beta_{n,t}$ happen to have higher $\sigma_n(d_{n,t})$, then we may still spuriously find price multipliers that decrease with shock size.

Under this alternative hypothesis, scaling demand shocks by stock-specific volatility

$$d^{\text{std}}_{n,t} = d_{n,t}/\sigma_n(d_{n,t}),$$

Panel A: price impact regressions									
	Dependent variable: stock return $r_{n,t}$								
	BMI			FIT			OFI		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$M_{\lvert d_{n,t}\rvert<\sigma}$	1.77***	1.74***	1.65***	3.92***	3.80***	3.76***	3.13***	3.11***	3.15***
	(0.50)	(0.53)	(0.52)	(0.52)	(0.53)	(0.52)	(0.10)	(0.10)	(0.10)
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	1.05*	1.22*	1.11*	3.33***	3.27***	3.21***	2.03***	1.95***	1.91***
	(0.58)	(0.67)	(0.67)	(0.30)	(0.32)	(0.33)	(0.07)	(0.07)	(0.07)
$M_{\lvert d_{n,t}\rvert>2\sigma}$	0.69**	0.58	0.56	2.69***	2.52***	2.56***	0.91***	0.78***	0.58***
	(0.29)	(0.43)	(0.40)	(0.26)	(0.24)	(0.24)	(0.06)	(0.06)	(0.08)
Predictor controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Liquidity controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Interacted: predictors	N	Y	Y	N	Y	Y	N	Y	Y
Interacted: liquidity	N	N	Y	N	N	Y	N	N	Y
Obs	9,910	9,910	9,910	561,404	561,404	561,404	527,744	527,744	527,744
R^2	0.172	0.208	0.226	0.080	0.085	0.087	0.129	0.142	0.155
Marginal $R^2(d_{n,t})$	0.014	0.014	0.013	0.005	0.004	0.004	0.049	0.043	0.039
Panel B: Coefficient differences									
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]} - M_{\lvert d_{n,t}\rvert<\sigma}$	−0.72	−0.52	−0.55	−0.59*	−0.53	−0.56*	−1.10***	−1.16***	−1.24***
	(0.59)	(0.63)	(0.67)	(0.35)	(0.34)	(0.33)	(0.07)	(0.07)	(0.06)
$M_{\lvert d_{n,t}\rvert>2\sigma} - M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	−0.36	−0.64	−0.54	−0.64***	−0.75***	−0.65***	−1.12***	−1.18***	−1.33***
	(0.54)	(0.84)	(0.81)	(0.24)	(0.23)	(0.24)	(0.05)	(0.05)	(0.06)
$M_{\lvert d_{n,t}\rvert>2\sigma} - M_{\lvert d_{n,t}\rvert<\sigma}$	−1.08**	−1.16*	−1.09*	−1.23***	−1.28***	−1.20***	−2.22***	−2.34***	−2.57***
	(0.46)	(0.69)	(0.64)	(0.44)	(0.44)	(0.44)	(0.08)	(0.07)	(0.08)

Table 3. Price multipliers controlling for demand-characteristic interactions

Panel A reports results for Fama-MacBeth regressions (6):

$$r_{n,t} = \sum_{\text{bin } b} M_{b,t} \cdot I_{|d_{n,t}| \in b} \cdot d_{n,t} + \boldsymbol{b}_t^{'} \cdot \boldsymbol{z}_{n,t-1} \cdot d_{n,t} + \boldsymbol{c}_t^{'} \boldsymbol{x}_{n,t-1} + \tau_t + \epsilon_{n,t} \tag{8}$$

where $r_{n,t}$ is stock return and $d_{n,t}$ represent demand shocks. In each time period, we split the sample into three bins b by comparing $|d_{n,t}|$ against the cross-sectional standard deviation of $d_{n,t}$. Columns (1) through (3) report results using the ΔBMI demand shock and monthly returns. Columns (4) through (6) and (7) through (9) report results using quarterly returns and the FIT and OFI demand shocks, respectively. Columns (1), (4), and (7) include the full set of controls in Table 2. Columns (2), (5), and (8) also control for demand interacted with 13 return-predicting stock characteristics; columns (3), (6), and (9) further control for demand interacted with 6 liquidity proxies. Those controls are described in Section 3.1. Panel B reports pairwise differences in price multipliers. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

should remove the variation in multipliers by shock size. Therefore, we split observations into bins using the standard deviation of $d_{n,t}^{\text{std}}$ which ensures that, approximately, each stock has the same fraction of observations in each bin. This new specification neutralizes the concern that our results are driven by compositional differences across bins.

We test this new specification in Appendix Table B.4. For each stock n in period t, we use the previous $h = 4, 8$, or 12 quarters to estimate $\sigma(d_{n,t})$ and construct standardized demand

shocks $d^{\text{std}}_{n,t}$ for FIT and OFI.[11] When using standardized demand shocks in regression (2), we continue to find similar evidence that multipliers decrease with shock size, which assuages the concern that unobserved variation in stock-specific multipliers drives our results.

Differential anticipation of demand shocks. One may be concerned that our results are driven by investors anticipating larger demand shocks to a greater extent than smaller shocks. Since anticipation can lead prices to change in advance of the actual occurrence of the shock (Gabaix and Koijen, 2022), this situation can cause large shocks to have a greater fraction of their price impact materialize *in advance*, and thus not be captured by our *contemporaneous* price impact estimates. To address this concern, we re-estimate regression (2) using cumulative returns from period $t-h$ to t ($r_{n,t-h\to t}$) as the dependent variable, thereby capturing pre-shock price drifts that occur in anticipation of the shock. Appendix Table B.5 shows that we continue to see multipliers declining with shock size after including pre-drifts, which assuages this concern.

Differential correlation with other investors. One may be concerned that our demand shocks correlate with the trading of other investors that we do not observe. In this case, our estimated price multipliers would be biased upward (i.e., a given price change would be ascribed to a too-small demand shock). Thus, if this correlation is stronger when our demand shocks are smaller, this may explain our results.

The concern about unobserved trading flows is a generic issue inherent with reduced-form studies of price impact because it is difficult to capture all demand variation. However, we argue that this alternative hypothesis is unlikely to explain our results.

We empirically address this concern in ways specific to our demand shocks. For BMI, because the Russell index reconstitutions are exogenous, the main concern is a differential "pass-through rate" of BMI changes to the total change of institutional stock ownership.

[11] We do not conduct the same exercise on BMI because it only realizes once a year, and so we cannot estimate demand shock volatility with precision.

Specifically, one may be worried that low (high) magnitude BMI changes have a higher (lower) pass-through rate. However, Appendix Table B.1 shows that the pass-through rate does not vary with the magnitude of BMI changes. For FIT, the concern is that mutual fund flows and other investors' trades correlate since they are partially driven by common factors. To guard against this correlation varying with flow size, we construct an alternate version of FIT where we remove common factors in mutual fund flows (in Appendix B.1). The price multiplier results based on this alternative FIT measure are qualitatively unchanged.[12]

Differential speed of return reversal. One may be concerned that the price impact of larger shocks, for some reason, reverts more quickly. If so, the change in expected return per unit of the demand shock would be larger for larger shocks. In this case, investors may respond more elastically to larger shocks, leading to smaller multipliers (as suggested theoretically by van Binsbergen et al. (2025)). We do not find evidence of this pattern. Appendix Table B.6 shows the price impact of larger shocks does not revert more quickly.

Differential measurement error. One may be concerned the multiplier variation we document arises from greater measurement error for larger demand shocks. Measurement error in the independent variable causes attenuation bias in the coefficient estimates. Therefore, if measurement error is larger when shocks are larger, we could spuriously find that price multipliers "decrease with shock size" due to greater attenuation bias for larger shocks.

While measurement error is potentially a concern for the contemporaneous regressions in Section 4.1, it is arguably less relevant for dynamic regressions in Section 4.3. We find that as cumulative *past* shocks grow in magnitude, the *contemporaneous* multiplier decreases. If measurement error is not serially correlated, attenuation bias cannot explain this finding.

It is also worth pointing out that measurement errors would have to be large to explain our

[12] We cannot conduct a similar test for OFI (which already captures all aggressive trades in stocks) because we do not have ways to capture demand shifts from non-aggressive trades, midpoint dark pool trades, or other trades not detected by the Lee-Ready algorithm.

findings. As a concrete example, consider the BMI-based estimates in column (3) of Table 2. The price multiplier for shocks in the largest bin is $1 - 0.69/1.77 \approx 61\%$ smaller than that in the smallest bin. Even if we assume there is no measurement error in the smallest bin, in order for this difference to be explained by attenuation bias, the standard deviation of the measurement error in the largest bin must be $\sqrt{\frac{0.61}{1-0.61}} \approx 1.25$ times as large as that of the true BMI variation.[13] If this were true, we should expect the pass-through rate from BMI to institutional ownership changes to be much lower when BMI has large magnitudes due to attenuation bias. However, this hypothesis is not supported in Appendix Table B.1.

4.3 Price Multipliers Increase with Cumulative Past Shock Size

The results in Section 4.1 suggest that multipliers decrease with shock size because larger shocks create larger price dislocations that expand market absorption capacity. This logic implies that multipliers should also decrease as *past* shocks accumulate and create larger dislocations. This section tests this implication.

Empirical test. We test how multipliers vary with the size of cumulative past shocks using FIT and OFI. We do not use BMI because it does not have consecutive realizations since it only features shocks in the month of June. We estimate this cross-sectional regression:

$$r_{n,t} = \sum_{\text{bin } b} M_{b,t} \cdot I_{\left|\sum_{l=1}^{L} d_{n,t-l}\right| \in b} \cdot d_{n,t} + \boldsymbol{c}_t' \boldsymbol{x}_{n,t-1} + \tau_t + \epsilon_{n,t}. \tag{9}$$

We form bins based on the magnitude of cumulative demand shocks over the past L quar-

[13] Suppose the true relationship is $r_{n,t} = a + b \cdot d_{n,t} + \epsilon_{n,t}$, but we mismeasure demand as $\tilde{d}_{n,t} = d_{n,t} + u_{n,t}$ where $u_{n,t}$ is mean-zero, normally distributed, classical measurement error. We can infer the required measurement error variation using the attenuation factor

$$\hat{b} = \frac{Cov(r_{n,t}, \tilde{d}_{n,t})}{Var(\tilde{d}_{n,t})} = \left(1 - \underbrace{\frac{\sigma_u^2}{\sigma_d^2 + \sigma_u^2}}_{\text{attenuation factor}}\right) \cdot b \to \sigma_u = \sqrt{\frac{\text{attenuation factor}}{1 - \text{attenuation factor}}} \cdot \sigma_d$$

ters ($\left|\sum_{l=1}^{L} d_{n,t-l}\right|$): $b \in \{[0, \sigma_t), [\sigma_t, 2\sigma_t], [2\sigma_t, \infty)\}$, where σ_t is the cross-sectional standard deviation of $\sum_{l=1}^{L} d_{n,t-l}$. We use $L = 4$ — one year of cumulative shocks — as the baseline. As in Section 4.1, we estimate the regression in each cross-section and report the time-series average coefficients, as in Fama and MacBeth (1973). We use the same controls as in Table 2.

Table 4 reports the results. Columns (1)–(4) and columns (5)–(8) report results for FIT and OFI, respectively; the different specifications vary the controls used (described in Tables 2 and 3). All specifications indicate that multipliers decline with cumulative past shock size.

Appendix Table B.7 shows that our results are robust to using alternative lookback windows to calculate cumulative past shock size. In Appendix Table B.8, we measure how multipliers vary with both contemporaneous and lagged shocks sizes. The results are broadly consistent with both having independent explanatory power over multiplier variation.

Overall, these results suggest that as demand shocks accumulate and create larger price dislocations, investors endogenously become more elastic, which expands market absorption capacity and lowers multipliers.

Other implications. The dynamic results in Table 4 help alleviate potential concerns about dynamic consistency. If price impact curves do not vary over time, but exhibit concavity in each period (as in Section 4.1), then one may be able to make profits through manipulative trades. Specifically, one can make profits by first buying 1 unit for N periods and then selling N units in period $N + 1$. While this is not a riskless arbitrage (it requires taking risk over $N + 1$ quarters), it would nonetheless be counterintuitive for such strategies to be profitable on average. Addressing this puzzle, Huberman and Stanzl (2004) theoretically predict that, if contemporaneous price impact curves are not linear, then price impacts should vary dynamically. Our results are consistent with their prediction.

Second, these dynamic findings also help alleviate measurement error concerns. As explained in Section 4.2, one may worry that larger shocks have greater measurement error, which cre-

Panel A: price impact regressions								
	Dependent variable: stock return $r_{n,t}$							
$d_{n,t}$ =	FIT				OFI			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert<\sigma}$	3.86***	3.72***	3.65***	3.64***	1.65***	1.73***	1.73***	1.84***
	(0.34)	(0.32)	(0.32)	(0.32)	(0.13)	(0.14)	(0.14)	(0.15)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert\in[\sigma,2\sigma]}$	2.79***	2.80***	2.71***	2.74***	1.21***	1.34***	1.25***	1.45***
	(0.33)	(0.33)	(0.32)	(0.32)	(0.12)	(0.12)	(0.11)	(0.12)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert>2\sigma}$	2.04***	2.01***	1.81***	1.89***	0.93***	1.04***	0.95***	1.20***
	(0.31)	(0.31)	(0.28)	(0.29)	(0.14)	(0.13)	(0.10)	(0.13)
Predictor controls	Y	Y	Y	Y	Y	Y	Y	Y
Liquidity controls	N	Y	Y	Y	N	Y	Y	Y
Interacted: predictors	N	N	Y	Y	N	N	Y	Y
Interacted: liquidity	N	N	N	Y	N	N	N	Y
Obs	478,324	478,324	478,324	478,324	423,696	423,696	423,696	423,696
R^2	0.070	0.085	0.091	0.094	0.108	0.127	0.142	0.154
Panel B: Coefficient differences								
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert\in[\sigma,2\sigma]} - M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert<\sigma}$	−1.06***	−0.93***	−0.94***	−0.90***	−0.43***	−0.39***	−0.48***	−0.39***
	(0.27)	(0.27)	(0.26)	(0.25)	(0.10)	(0.09)	(0.08)	(0.05)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert>2\sigma} - M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert\in[\sigma,2\sigma]}$	−0.75***	−0.78***	−0.91***	−0.85***	−0.29***	−0.30***	−0.29***	−0.25***
	(0.25)	(0.25)	(0.22)	(0.23)	(0.08)	(0.08)	(0.07)	(0.06)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert>2\sigma} - M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert<\sigma}$	−1.82***	−1.71***	−1.85***	−1.75***	−0.72***	−0.69***	−0.77***	−0.64***
	(0.35)	(0.34)	(0.31)	(0.31)	(0.16)	(0.15)	(0.12)	(0.08)

Table 4. Price multipliers by cumulative past demand shock sizes

For each demand shock (FIT and OFI), we estimate Fama-MacBeth regressions:

$$r_{n,t} = \sum_{\text{bin } b} M_{b,t} \cdot I_{\left|\sum_{l=1}^{L} d_{n,t-l}\right| \in b} \cdot d_{n,t} + \boldsymbol{c}_t^{'} \boldsymbol{x}_{n,t-1} + \tau_t + \epsilon_{n,t} \tag{11}$$

where $r_{n,t}$ is stock return, $d_{n,t}$ is the demand shock, and $L = 4$ is the lookback window (number of quarters). In each period, we split the sample into three bins b by comparing $\left|\sum_{l=1}^{L} d_{n,t-l}\right|$ with σ which is defined as the cross-sectional standard deviation of $\sum_{l=1}^{L} d_{n,t-l}$. The first four columns use FIT as the demand shock and the last four columns use OFI. Columns for each demand shock differ by the controls used. Panel A reports regression estimates. Panel B reports pairwise coefficient differences. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

ates more attenuation bias that spuriously reduces their estimated multipliers. However, absent serial correlation in measurement error, attenuation bias cannot explain why multipliers decrease with the magnitude of *past* demand shocks.

5 Evidence from Price Elasticities of Demand

The inverse relationship between price multipliers and demand shock size documented in Section 4 suggests investors become more willing to absorb shocks as price dislocations — and so profit opportunities — grow. This section tests this interpretation directly in

investor holdings data. We find investors become more price elastic in response to larger price dislocations. That is, investors require less per-unit compensation to absorb shocks when price dislocations are larger. This behavior expands market absorption capacity following large shocks. Section 5.1 describes our empirical approach. Section 5.2 presents our results.

5.1 Empirical Approach

While the demand shocks in Section 4 generate enough price variation to estimate price multipliers, they do not generate enough variation to precisely estimate investor-specific price elasticities, given the large degree of variation in holdings data. Hence, we need an shock with more variation. To this end, we estimate elasticities using the granular instrumental variable (GIV) methodology of Gabaix and Koijen (2022, 2024). Specifically, we develop a dynamic, nonlinear extension of the optimal GIV methodology from Chaudhary et al. (2024), which is designed to maximize statistical power when estimating heterogeneous elasticities.

Section 5.1.1 describes how we model asset demand. Section 5.1.2 explains how the GIV methodology identifies elasticities. Section 5.1.3 discusses implementation details.

5.1.1 Nonlinear Model for Investor Demand

Section 4.3 reveals that price multipliers decline as cumulative demand shocks grow in magnitude. Motivated by this observation, we model investors' asset demand as:

$$\Delta q_{i,n,t} = -\underbrace{\left(\zeta_{1,i,t} + \zeta_{2,i,t}\tilde{P}_{n,t}\right)}_{\text{Price Elasticity of Demand}} \Delta p_{n,t} + \underbrace{\boldsymbol{\lambda}_{i,t}^{'}\boldsymbol{\eta}_{n,t} + u_{i,n,t}}_{\text{Demand Shock}} \tag{12}$$

$$\tilde{P}_{n,t} = \left|\sum_{l=1}^{L} \Delta p_{n,t-l}\right| - \mathbb{E}_t^{CX}\left[\left|\sum_{l=1}^{L} \Delta p_{n,t-l}\right|\right].$$

where $\Delta q_{i,n,t}$ is the percentage change in quantity of shares demanded by investor i for stock n in quarter t. $\mathbb{E}_t^{CX}$ denotes the cross-sectional average across stocks n in quarter t. Appendix C.1 shows that the functional form of this asset demand specification can generate the price dynamics we find in Section 4.3.

In equation (12), the first component of the quantity change captures movement along the demand curve: how demand responds to percentage price changes ($\Delta p_{n,t}$), holding the demand shock fixed. The second component captures demand shocks (i.e., demand curve shifts): changes in non-price determinants of demand, such as news about cash flows or risk.

The demand shock has two components. First, there is a common component $\boldsymbol{\eta}_{n,t}$ that affects all investors' demand for stock n (with heterogeneous loadings $\boldsymbol{\lambda}_{i,t}$). This common shock includes stock-quarter specific characteristics such as firm-level fundamentals or news. Second, there is an uncorrelated residual demand shock $u_{i,n,t}$ that is assumed to be conditionally mean-independent across investors, within and across quarters:

$$\forall i \neq j, \forall t, \forall l = 0, \ldots, L : \mathbb{E}^{CX}[u_{i,n,t} \mid u_{j,n,t-l}, \boldsymbol{\eta}_{n,t}] = \mathbf{0}. \tag{13}$$

This residual shock captures investors' idiosyncratic beliefs or preferences for stock n.

Our main goal is to test if $\zeta_{2,i,t} > 0$. $\zeta_{1,i,t}$ represents the average elasticity, while $\zeta_{2,i,t}$ captures how elasticity changes with the (demeaned) magnitude of cumulative price changes over the past L quarters ($\tilde{P}_{n,t}$), *holding the demand shock fixed*. Thus, finding $\zeta_{2,i,t} > 0$ ($\zeta_{2,i,t} < 0$) indicates that demand becomes more (less) elastic as the magnitude of *exogenous* cumulative past price changes (i.e., price dislocations) grows.

Under the assumption of homogeneous substitution conditional on observables, $\zeta_{1,i,t} + \zeta_{2,i,t}\tilde{P}_{n,t}$ is the "relative" (Haddad et al. (2025)) or "micro" elasticity (Gabaix and Koijen, 2022).[14]

5.1.2 Identifying Demand Elasticities with GIV

The challenge in identifying elasticities is that quantities and prices are simultaneously determined in equilibrium. Through market clearing, investor i's demand shock correlates with the contemporaneous price change. Furthermore, if demand shocks are serially correlated,

[14]The relative elasticity is the difference between the own-price and cross-price elasticities. It measures the demand response for an asset relative to another with identical characteristics when relative prices change.

i's current demand shock also correlates with lagged price changes. Thus, a regression of quantity changes on current and lagged price changes fails to identify $\zeta_{1,i,t}$ and $\zeta_{2,i,t}$ due to omitted variable bias. Instead, we need exogenous variation in current and lagged price changes that is uncorrelated with investor i's demand shocks.

To address the simultaneity problem, we use the optimal granular instrumental variables (GIV) methodology of Chaudhary et al. (2024). At a high level, this method uses the residual demand shocks of other investors $\{u_{j,n,t}\}_{j\neq i}$ to instrument the price changes faced by each investor i. In particular, to accommodate the dependence of the elasticity on lagged price changes, we develop a dynamic, nonlinear extension of this method.

To convey the intuition, we derive the price change due to all investors' demand shocks. Assuming a fixed supply of shares, the aggregate equilibrium change in quantity demanded must be zero: $0 = \sum_i S_{i,n,t-1}\Delta q_{i,n,t}$. $S_{i,n,t-1}$ are lagged ownership share weights: the proportion of outstanding shares held by investor i. After plugging (12) into this market-clearing expression, a first-order approximation around $\tilde{P}_{n,t} = 0$ yields the equilibrium price change:

$$\Delta p_{n,t} \approx \underbrace{\left(\frac{1}{\zeta_{1,S,t}} - \frac{\zeta_{2,S,t}}{\zeta_{1,S,t}^2}\tilde{P}_{n,t}\right)}_{\text{Price Multiplier}} \left(\boldsymbol{\lambda}_{S,t}'\boldsymbol{\eta}_{n,t} + u_{S,n,t}\right). \tag{14}$$

The S subscript denotes ownership-share weighted averages (e.g., $u_{S,n,t} = \sum_i S_{i,n,t-1}u_{i,n,t}$). Intuitively, the demand shocks of investors with larger ownership shares have a greater impact on prices. Moreover, the price multiplier decreases with the weighted-average unconditional price elasticity $\zeta_{1,S,t}$. Crucially, the multiplier can depend on the magnitude of exogenous past price changes ($\tilde{P}_{n,t}$), consistent with our findings in Section 4.3. If investors become more elastic as exogenous past price changes grow (i.e., if the weighted-average $\zeta_{2,S,t} > 0$), then the multiplier decreases with the magnitude of exogenous past price changes.

Equation (14) implies that we can instrument for the price change faced by investor i using

the residual demand shocks of *other* investors. Since residual demand shocks are assumed to be uncorrelated across investors (by (13)), any linear combination of residual demand shocks $u_{j,n,t}$ for other investors $j \neq i$ is uncorrelated with i's demand shock $\boldsymbol{\lambda}_{i,t}^{'}\boldsymbol{\eta}_{n,t} + u_{i,n,t}$. Moreover, (14) implies that the linear combination that weights $u_{j,n,t}$ by ownership shares will be maximally correlated with price, and thus provide the most relevant price instrument (Chaudhary et al. (2024)). Therefore, we form the following granular instrumental variable (GIV) for the price change investor i faces for stock n:

$$z_{i,n,t} = \sum_{j \neq i} S_{j,n,t-1} u_{j,n,t}. \tag{15}$$

$z_{i,n,t}$ provides exogenous variation in $\Delta p_{n,t}$ uncorrelated with i's demand shock.

Given this GIV, the following two moment conditions identify $\zeta_{1,i,t}$ and $\zeta_{2,i,t}$ for each i, t:

$$0 = \mathbb{E}_t^{CX}\left[u_{i,n,t} \cdot z_{i,n,t} \mid \boldsymbol{\eta}_{n,t}\right] \tag{16}$$

$$0 = \mathbb{E}_t^{CX}\left[u_{i,n,t} \cdot z_{i,n,t} \cdot \tilde{Z}_{i,n,t} \mid \boldsymbol{\eta}_{n,t}\right], \tag{17}$$

$$\tilde{Z}_{i,n,t} = \left|\sum_{l=1}^{L} z_{i,n,t-l}\right| - \mathbb{E}_t^{CX}\left[\left|\sum_{l=1}^{L} z_{i,n,t-l}\right|\right]$$

where the expectations are taken across stocks n in quarter t. The moment conditions above are formally equivalent to a two-stage least squares (2SLS) regression of $\Delta q_{i,n,t}$ on both $\Delta p_{n,t}$ and the interaction term $\Delta p_{n,t} \cdot \tilde{P}_{n,t}$, with $z_{i,n,t}$ and $z_{i,n,t} \cdot \tilde{Z}_{i,n,t}$ as instruments, and $\boldsymbol{\eta}_{n,t}$ included as controls. $z_{i,n,t}$ and $\tilde{Z}_{i,n,t}$ provide exogenous variation in the current price change and the magnitude of past price changes.

Crucially, to observe other investors' residual demand shocks and construct these instruments, we must know other investors' elasticity parameters $\{\zeta_{1,j,t}, \zeta_{2,j,t}\}_{j \neq i}$:

$$u_{j,n,t} = \Delta q_{j,n,t} + \left(\zeta_{1,j,t} + \zeta_{2,j,t}\tilde{P}_{n,t}\right)\Delta p_{n,t} - \boldsymbol{\lambda}_{j,t}^{'}\boldsymbol{\eta}_{n,t},$$

For this reason, we estimate $\{\zeta_{1,i,t}, \zeta_{2,i,t}\}_i$ for all investors simultaneously using generalized method of moments (GMM) with moment conditions (16) and (17).

5.1.3 Empirical Implementation

We estimate demand curve (12) for all investors in the SEC Form 13F institutional holdings data.[15] We restrict our sample to intensive-margin quantity changes, defined as observations where investor i holds a positive quantity of stock n in both quarters $t-1$ and t.

While we seek to characterize heterogeneity in the elasticity parameters across institutions, we lack the statistical power to flexibly estimate these parameters at the institution level. Thus, we parsimoniously capture this heterogeneity by parameterizing $\zeta_{1,i,t}$ and $\zeta_{2,i,t}$ as affine functions of institution i's demeaned active share:

$$\zeta_{k,i,t} = \zeta_{k,0,t} + \zeta_{k,\text{Active Share},t} \cdot \text{Active Share}_{i,t-1-L}, \qquad \text{for } k = 1, 2. \tag{18}$$

Motivated by Cremers and Petajisto (2009), the active share for institution i is defined as $\frac{\sum_n |w_{i,n,t} - w_{i,n,t}^{\text{benchmark}}|}{2}$, where $w_{i,n,t}$ and $w_{i,n,t}^{\text{benchmark}}$ are the actual and benchmark portfolio weights. We proxy for the latter using the market cap weights for each institution's investment universe.[16] In (18), we use the four-quarter rolling average active share from quarter $t-4-L$ to $t-1-L$ to avoid correlation with demand shocks from quarters $t-L$ to t. Moreover, we demean the active share measure within each quarter so $\zeta_{k,0,t}$ represents the $\zeta_{k,i,t}$ parameter for the average institution. This parametric approach preserves statistical

[15] Following Davis and Haltiwanger (1992), we calculate $\Delta q_{i,n,t}$ as:

$$\Delta q_{i,n,t} = \frac{\hat{Q}_{i,n,t} - \hat{Q}_{i,n,t-1}}{\frac{1}{2}(\hat{Q}_{i,n,t} + \hat{Q}_{i,n,t-1})},$$

where $\hat{Q}_{i,n,t-1} = H_{i,n,t-1}$ is investor i's dollar holdings in stock n in quarter $t-1$, and $\hat{Q}_{i,n,t} = H_{i,n,t}/(1 + R^X_{n,t-1\to t})$ is the dollar holdings in quarter t adjusted for the ex-dividend return $R^X_{n,t-1\to t}$. This specification is motivated by potential measurement errors in share data (e.g., unadjusted stock splits) in 13F filings. Using dollar holdings circumvents this issue, while the denominator, which maps the expression into the range $[-2, 2]$, effectively winsorizes large percentage changes.

[16] Following Koijen and Yogo (2019), an institution's universe is all stocks held in the past 11 quarters.

power while allowing us to characterize heterogeneity across institutions.

We estimate the elasticity parameters in the cross-section of stocks in each quarter. We use $L = 4$ quarters to compute the cumulative lagged price change, as in Section 4.3. To improve numerical stability, we perform the estimation in four-quarter rolling windows and report the average parameter estimates across all windows. We compute Fama-MacBeth standard errors with a Newey-West correction (8 lags) to account for the overlapping windows.

The stock characteristic controls, $\boldsymbol{\eta}_{n,t}$, include the return-predicting characteristics discussed in Section 3.1, dummies for the Fama-French twelve industries, and five latent characteristics estimated via principal component analysis on the within-quarter institution $\times$ stock $\Delta q_{i,n,t}$ panel.[17] We also include institution-quarter fixed effects to ensure the elasticity parameters are identified solely from cross-sectional variation within each institution's portfolio. Appendix C.2 provides further estimation details.

Table 5 summarizes the quarterly 13F holdings data from 1982 to 2021.[18] The mean $\Delta q_{i,n,t}$ is around zero, and one standard deviation is 41%, which corresponds to buying or selling 41% of pre-existing holdings. The absolute value of the cumulative price change in the past $L = 4$ quarters has an average of 30% and a standard deviation of 33%. The mean active share is 0.41 and the standard deviation is 0.20.

5.2 Empirical Results

Table 6 displays the elasticity parameter estimates. The first two columns display the time-series average estimates for $\zeta_{1,0}$ and $\zeta_{1,\text{Active Share}}$, which control the elasticity for the average stock (i.e., when the demeaned absolute cumulative price change $\tilde{P}_{n,t} = 0$). We find $\zeta_{1,0} = 0.49$: on average institutions reduce their quantity demanded by 0.49% in response to an exogenous 1% price increase.[19] We find a positive $\zeta_{1,\text{Active Share}}$: more active institutions

[17] Our results are robust to using different numbers of latent characteristics, as shown in Appendix C.3.

[18] The data starts in 1980, but our sample begins after two years due to the four-quarter lookback window for $\tilde{P}_{n,t}$ and the four-quarter rolling estimation window.

[19] This estimate is similar to the institution-level elasticity estimates from Koijen and Yogo (2019).

	Obs	Mean	StDev	Percentiles 1%	5%	25%	50%	75%	95%	99%
$\Delta q_{i,n,t}$	24,671,308	-0.02	0.41	-1.57	-0.66	-0.07	-0.00	0.06	0.55	1.46
$\Delta p_{n,t}$	24,671,308	0.04	0.18	-0.44	-0.25	-0.06	0.03	0.13	0.32	0.56
$\left\vert\sum_{l=1}^{L} \Delta p_{n,t-l}\right\vert$	24,671,308	0.30	0.33	0.00	0.02	0.10	0.21	0.39	0.81	1.50
Active Share$_{i,t}$	24,671,308	0.41	0.20	0.07	0.09	0.27	0.42	0.55	0.75	0.87
$z_{i,n,t}$	24,671,308	-0.02	0.06	-0.21	-0.11	-0.04	-0.02	0.01	0.06	0.12

Table 5. Summary statistics for price elasticity estimation

This table provides summary statistics for the data underlying the estimation of demand curve (12) in Section 5. Variables include quarterly percentage changes in quantity of shares demanded ($\Delta q_{i,n,t}$), quarterly percentage changes in price ($\Delta p_{n,t}$), absolute cumulative percentage price changes over the past four quarters ($|\sum_{l=1}^{L=4} \Delta p_{n,t-l}|$), institution-level active share (Active Share$_{i,t}$), and estimated GIV ($z_{i,n,t}$) defined as in (15). All variables except active share are expressed in absolute terms (i.e., 0.01 is 1%). The sample for all variables consists of quarterly data from 1982 to 2021.

have higher elasticities. The magnitude of $\zeta_{1,\text{Active Share}} = 1.58$ impies that a one standard deviation increase in active share above the mean raises elasticity by about $1.58 \times 0.2 = 0.32$, or from $\zeta_{1,0} = 0.49$ for the average institution to 0.81, an economically significant increase.

	$\zeta_{1,0}$	$\zeta_{1,\text{Active Share}}$	$\zeta_{2,0}$	$\zeta_{2,\text{Active Share}}$
Coefficient	0.49***	1.58***	0.57**	2.68**
	(0.05)	(0.22)	(0.28)	(1.10)

Table 6. Demand elasticity parameter estimates

We estimate demand curve (12) using quarterly 13F holdings data in rolling annual windows and report the average coefficient values over all time periods. We report Fama-MacBeth standard errors with a Newey-West correction of 8 lags. The first column ($\zeta_{1,0}$) displays the average elasticity for institutions with average active share. The second column ($\zeta_{1,\text{Active Share}}$) displays how much the average elasticity varies with active share. The third column ($\zeta_{2,0}$) displays how much elasticity changes with the magnitude of past exogenous cumulative price changes for the average institution (i.e., a 1% higher cumulative price change raises elasticity by $\zeta_{2,0} \cdot 0.01$). The fourth column ($\zeta_{2,\text{Active Share}}$) displays how much this sensitivity of elasticity to past exogenous cumulative price changes varies with active share. We control for the return-predicting characteristics from Table 2, dummies for the Fama-French twelve industries, and five latent stock-characteristics estimated using PCA. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

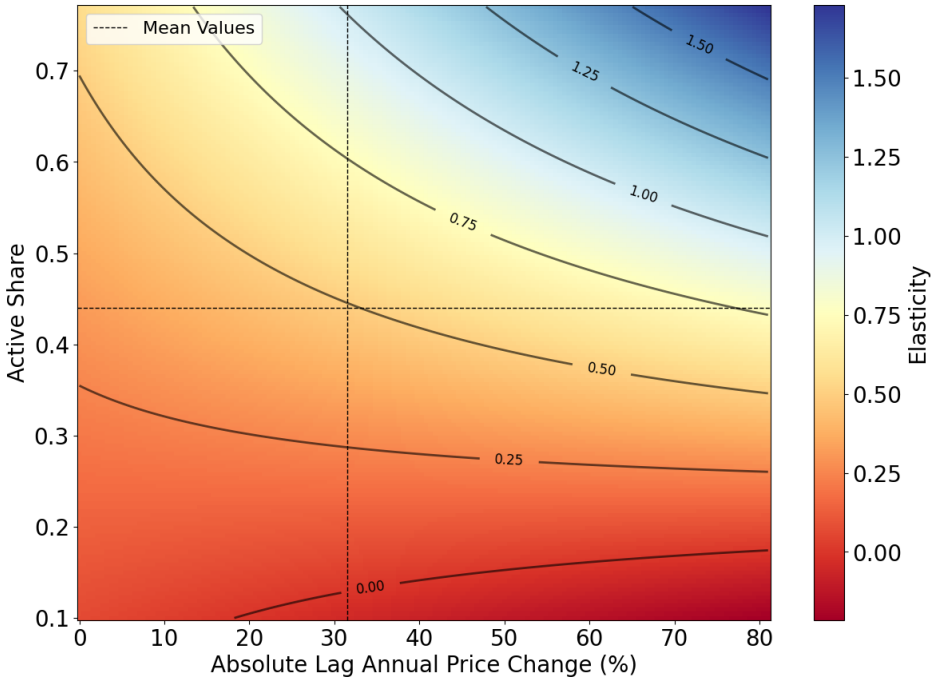
The last two columns display the estimates of $\zeta_{2,0}$ and $\zeta_{2,\text{Active Share}}$ — the parameters of interest that control how the elasticity varies with the magnitude of price dislocations. In the third column, we find $\zeta_{2,0} > 0$: the average institution becomes more elastic as exogenous cumulative past price changes grow in magnitude. In the fourth column we find $\zeta_{2,\text{Active Share}} > 0$, which implies that more active institutions become even more elastic as

exogenous cumulative past price changes grow in magnitude.

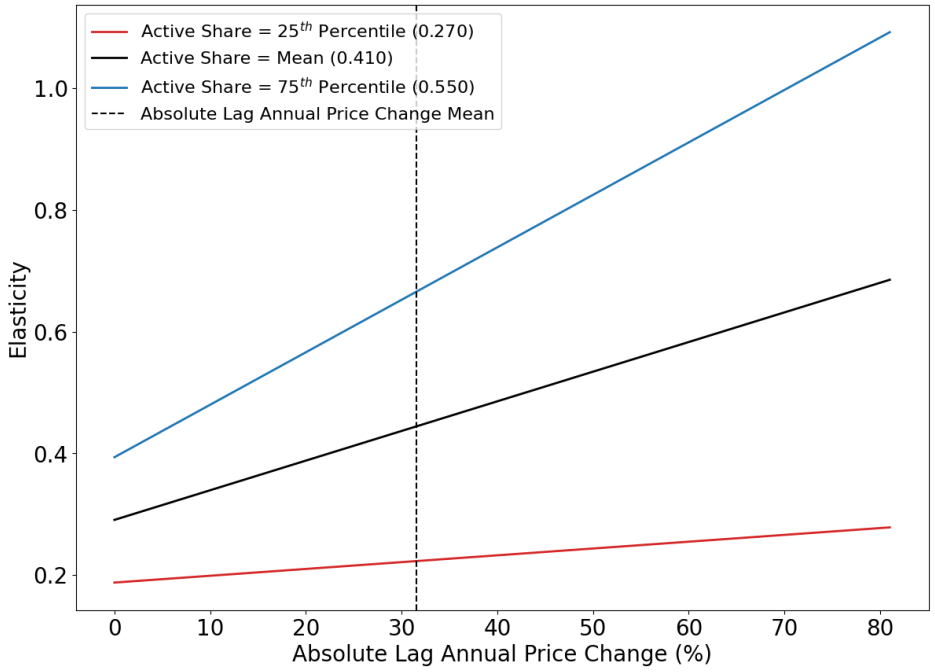
The magnitudes of $\zeta_{2,0}$ and $\zeta_{2,\text{Active Share}}$ are economically significant. Figure 2 Panel (a) displays total elasticity as a function of active share and absolute cumulative price change. Figure 2 Panel (b) displays total elasticity as a function of absolute cumulative price change for fixed active shares. For the average institution, moving from the 5th to the 95th percentile of absolute cumulative price change (from approximately 2% to 81% from Table 5) raises the total elasticity from 0.29 to 0.69. For more active institutions in the 75th percentile of active share, total elasticity rises from 0.39 to 1.09. Even for relatively inactive institutions in the 25th percentile of active share, the total elasticity rises from 0.19 to 0.28.

Appendix C.3.1 presents elasticity estimates when controlling for observed characteristics and up to ten latent characteristics. Controlling for alternative numbers of latent factors yields quantitatively similar (and statistically indistinguishable) parameter estimates.

Overall, we find most investors become more elastic following larger price dislocations. Thus, investors become more willing to absorb shocks as price dislocations grow. This behavior expands market absorption capacity following large shocks, and so lowers price multipliers.



(a) Elasticity as a Function of Active Share and Past Price Changes



(b) Elasticity as a Function of Past Price Changes for Fixed Active Shares

Figure 2. Visualizing variation in price elasticities

We visualize the variation in institutional price elasticities estimated in Section 5.2. Panel (a) uses a heatmap to show the estimated elasticities as a function of the absolute lagged four-quarter price change ($|\sum_{l=1}^{L=4} \Delta p_{n,t-l}|$) on the x-axis and active share on the y-axis. Panel (b) plots the estimated elasticities as a function of past price changes for different fixed active shares.

6 Mechanisms

Our empirical results in Sections 4 and 5 are consistent with mechanisms in which investors endogenously become more price elastic — more willing to absorb shocks — when large shocks create large price dislocations and large profit opportunities. Here we present two stylized models to illustrate two such mechanisms: fixed adjustment costs and endogenous inattention. Both mechanisms likely operate to some extent, and other forces may also contribute to the endogenous elasticity and multiplier behavior we document. We leave a quantitative decomposition of the relative importance of these mechanisms to future work.

6.1 Fixed Adjustment Costs

The first model we consider features a fixed cost that some investors face to adjust their holdings. This fixed cost can represent a physical transaction cost, a psychological cost of adjustment, or simply the administrative burden of having to organize an investment committee meeting to discuss an opportunity. Intuitively, these investors do not trade against small shocks that create only small price dislocations and small expected returns; the fixed cost outweights the small benefit. But as shocks grow and create larger price dislocations, expected returns become large enough to justify paying the fixed cost. As a result, for large enough price dislocations, investors endogenously become more elastic and multipliers fall.

The model has three periods: $t = 0, 1, 2$. At $t = 0$, investors receive a fixed endowment of shares. At $t = 1$, a supply shock occurs, the asset market clears, and the equilibrium price is determined. At $t = 2$, the asset's payoff is realized.

Assets. There is one risky asset that pays a dividend at $t = 2$: $\tilde{D} \sim N\left(\bar{D}, \sigma_D^2\right)$. All investors are endowed with Θ_0 shares of the asset at $t = 0$. At $t = 1$, an exogenous supply shock changes the total supply to Θ_1. This supply should be interpreted as the *residual* supply faced by optimizing investors: the total fixed supply less the demand from noise traders outside the model. The exogenous risk-free rate is normalized to zero.

Agents and preferences. There are two types of investors. First, there is mass $1-\epsilon$ of type-I (inelastic) investors who choose portfolios at $t=1$ to maximize mean-variance utility over $t=2$ terminal wealth subject to a fixed adjustment cost $C(Q_I)$ (as in Lo et al. (2004)):

$$\max_{Q_I} \quad Q_I \mathbb{E}[\tilde{D}-P] - \frac{\gamma}{2} Q_I^2 \mathbb{V}\left[\tilde{D}-P\right] - C\left(Q_I\right) \tag{19}$$

$$C(Q_I) = \begin{cases} 0 & , Q_I = \Theta_0 \\ \lambda & , Q_I \neq \Theta_0 \end{cases} \tag{20}$$

Second, there is a small mass $0<\epsilon<1$ of type-E (elastic) investors without the cost:

$$\max_{Q_E} \quad Q_E \mathbb{E}[\tilde{D}-P] - \frac{\gamma}{2} Q_E^2 \mathbb{V}\left[\tilde{D}-P\right]$$

The endogenous multiplier behavior arises from the type-I investors; these type-E investors simply ensure the market always clears, even if the type-I investors do not adjust.

Conditional on adjusting, all investors (both types I and E) choose:

$$Q = \frac{\mathbb{E}\left[\tilde{D}-P\right]}{\gamma \mathbb{V}\left[\tilde{D}\right]}. \tag{21}$$

However, as shown in Appendix D.1, the fixed cost gives rise to (S,s)-style adjustment behavior (Scarf, 1960). If the change in expected return (versus the expected return that prevails under $\Theta_1=\Theta_0$) is small enough in magnitude

$$\left|\mathbb{E}\left[\tilde{D}-P\right] - \Theta_0 \gamma \sigma_D^2\right| \leq \sqrt{2\lambda\gamma\sigma_D^2},$$

then the type-I investors do not adjust and choose $Q_I=\Theta_0$.

Equilibrium. Market-clearing implies

$$(1-\epsilon)Q_I + \epsilon Q_E = \Theta_1. \tag{22}$$

Appendix D.1 shows that substituting in the optimal Q_I and Q_E yields the equilibrium price:

$$P = \begin{cases} \bar{D} - \frac{\gamma\sigma_D^2(\Theta_1 - (1-\epsilon)\Theta_0)}{\epsilon} & , |\Theta_1 - \Theta_0| \leq \sqrt{2\lambda\gamma\sigma_D^2} \\ \bar{D} - \gamma\sigma_D^2\Theta_1 & , |\Theta_1 - \Theta_0| > \sqrt{2\lambda\gamma\sigma_D^2} \end{cases}.$$

Price multiplier. The price multiplier — the per-unit change in price in $t = 1$ due to the supply shock $\Theta_1 - \Theta_0$ — is (expressed as a positive number)[20]

$$M\left(\Theta_1 - \Theta_0\right) \equiv -\frac{P(\Theta_1) - P(\Theta_0)}{\Theta_1 - \Theta_0} = \begin{cases} \frac{\gamma\sigma_D^2}{\epsilon} & , |\Theta_1 - \Theta_0| \leq \frac{\sqrt{2\lambda}}{\sqrt{\gamma\sigma_D^2}} \\ \gamma\sigma_D^2 & , |\Theta_1 - \Theta_0| > \frac{\sqrt{2\lambda}}{\sqrt{\gamma\sigma_D^2}} \end{cases}$$

which is decreasing in the supply shock magnitude $|\Theta_1 - \Theta_0|$ since $\epsilon < 1$. Intuitively, for small shocks, the initial position of type-I investors is close to optimal, so the fixed cost exceeds the benefit of adjustment. Large shocks, however, significantly alter the optimal quantity (since they alter expected return), making adjustment worthwhile. Thus, for sufficiently large shocks, type-I investors endogenously become more elastic: they become more willing to absorb shocks, which expands market absorption capacity and lowers the multiplier.

6.2 Endogenous Inattention

The second model we consider features endogenous inattention. In general, investors face uncertainty about investment opportunities that renders their demand inelastic. However,

[20] Note that in this stylized model where portfolio choice is in the number of shares (instead of portfolio weights) and dollar expected returns (instead of percentage expected returns), this definition of price multiplier is the natural object for measuring price impact, as opposed to the definition we use empirically in terms of *percentage* price and quantity changes.

large shocks create large price dislocations and large expected returns, which incentivize investors to acquire more information and reduce their uncertainty. In practice, acquiring information can mean paying more attention to a stock, acquiring data or reports, organizing a meeting with company management for due diligence, or even hiring more analysts. As a result, investors endogenously become more elastic and multipliers fall.

The model has three periods: $t = 0, 1, 2$. Information choice occurs in $t = 0$. The asset market clears in both $t = 0$ and $t = 1$. Asset payoffs are realized in $t = 2$.

Assets. There is one asset that pays a risky dividend in period $t = 2$:

$$\tilde{D} = \bar{D} + \eta + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma_\varepsilon^2\right), \quad \eta \sim N\left(0, \sigma_\eta^2\right).$$

ε is unforecastable, but η can be learned about at a cost (explained below). The asset has exogenous supply of Θ_t in period $t \in \{0, 1\}$, with the time $t = 1$ change distributed as $\Theta_1 - \Theta_0 \sim N(0, \sigma_\Theta^2)$. The exogenous risk-free rate is normalized to zero.

Agents and preferences. A representative investor chooses her portfolio at $t = 0$ and 1 to maximize mean-variance utility over terminal wealth

$$\max_{Q_t} \mathbb{E}_t\left[Q_t\left(\tilde{D} - P\right)\right] - \frac{\gamma}{2}\mathbb{V}_t\left[Q_t\left(\tilde{D} - P\right)\right] \tag{23}$$

Information. At $t = 0$, the investor knows the asset pricing parameters $\bar{D}, \sigma_\varepsilon^2$, and σ_Θ^2. She does not know η, but has the objectively correct prior: $\eta \sim N\left(0, \sigma_\eta^2\right)$. After observing Θ_0, the investor can pay cost $C(G)$ at $t = 0$ to acquire a signal s for η at $t = 1$:[21]

$$\begin{aligned} s &= \eta + u, u \sim N\left(0, \sigma_u^2\right) \\ G &= \frac{\sigma_u^{-2}}{\sigma_u^{-2} + \sigma_\eta^{-2}}. \end{aligned} \tag{24}$$

[21] As in e.g., Kyle (1989); Van Nieuwerburgh and Veldkamp (2009, 2010); Kacperczyk et al. (2016).

where G denotes the Bayesian gain of signal s. The investor's posterior belief at $t = 1$ is

$$\eta \sim N\left(G \cdot s, (1 - G) \cdot \sigma_\eta^2\right).$$

Thus, a larger gain G corresponds to lower uncertainty (i.e., lower posterior variance). The marginal cost of reducing uncertainty is positive and increasing ($C^{'}(G) > 0, C^{''}(G) > 0$).

Since the investor does not know at $t = 0$ what signal s she will observe at $t = 1$, she chooses G to maximize *expected* utility at $t = 0$, integrating over all possible realizations of s:

$$\max_G \mathbb{E}_0\left[\mathbb{E}_1\left[Q_1\left(\tilde{D} - P\right)\right] - \frac{\gamma}{2}\mathbb{V}_1\left[Q_1\left(\tilde{D} - P\right)\right]\right] - C\left(G\right) \tag{25}$$

The investor's beliefs are rational: her $t = 0$ beliefs about future prices and quantities are consistent with the true equilibrium distributions.

Equilibrium definition. An equilibrium is defined as a set of portfolio choices (Q_t), information choice (G), and asset prices (P_t) such that:

1. Information choice G at $t = 0$ maximizes (25) given the investor's prior beliefs about η and her knowledge of Θ_0, the supply at $t = 0$.

2. Portfolio choices Q_t at $t = 0, 1$ maximize (23) given the investor's information.

3. The asset market clears in $t = 0, 1$:

$$Q_t = \Theta_t. \tag{26}$$

Equilibrium. We solve the model backward. First, we solve the investor's portfolio choice problem (23) in $t = 0$ and $t = 1$ given the information sets:

$$\begin{aligned} Q_0 &= \frac{\bar{D} - P_1}{\gamma\left(\sigma_\varepsilon^2 + \sigma_\eta^2\right)} \\ Q_1\left(\mathcal{I}_1\right) &= \frac{\bar{D} + G \cdot s - P_2}{\gamma\left(\sigma_\varepsilon^2 + (1 - G) \cdot \sigma_\eta^2\right)}. \end{aligned} \tag{27}$$

$\mathcal{I}_1 = (s, G)$ is the equilibrium $t = 1$ information set.

Second, we impose market clearing (26) to solve for the equilibrium prices as a function of the investor's information choice:

$$\begin{aligned} P_0 &= \bar{D} - \Theta_0 \gamma \left(\sigma_\varepsilon^2 + \sigma_\eta^2\right) \\ P_1 &= \bar{D} + G \cdot s - \Theta_1 \gamma \left(\sigma_\varepsilon^2 + \sigma_\eta^2 (1 - G)\right). \end{aligned} \tag{28}$$

Lastly, we solve for the investor's ex-ante optimal gain G at $t = 0$ that maximizes (25), which pins down the equilibrium price and quantity in both periods. Appendix D.2 shows the first-order condition for the ex-ante information choice problem (25) equates the marginal benefit and marginal cost of information acquisition:

$$\frac{\gamma}{2} \sigma_\eta^2 \left(\Theta_0^2 + \sigma_\Theta^2\right) = C'(G). \tag{29}$$

Since the cost C is convex in the chosen gain G, (29) implies that the optimal gain is increasing in the magnitude of $t = 0$ supply Θ_0. Intuitively, the marginal benefit of acquiring information and reducing uncertainty is larger when investors believe expected returns are higher: lower uncertainty allows them to take larger positions to more aggressively exploit the high expected returns without incurring excessive perceived risk. Since larger supply shocks create larger expected returns, the optimal gain increases with $|\Theta_0|$.

Price multiplier. Thus, from (28), the price multiplier in $t = 1$ is

$$M\left(\Theta_1 - \Theta_0\right) \equiv -\frac{P(\Theta_1) - P(\Theta_0)}{\Theta_1 - \Theta_0} = \gamma \left(\sigma_\varepsilon^2 + \sigma_\eta^2 (1 - G)\right).$$

Since the equilibrium gain G increases with the magnitude of Θ_0 by (29), the multiplier decreases with $|\Theta_0|$. Thus, the price multiplier is smaller for larger accumulated shocks, consistent with our dynamic multiplier findings in Section 4.3. This result is also as consistent with our static findings in Section 4.1 since one month or quarter is long enough for both information acquisition and trading to have occurred (i.e., one can think of $t = 0, 1$ as two

sub-periods within a month or quarter). Intuitively, large shocks create large price dislocations and large expected returns, which raise the marginal benefit of information acquisition. As a result, investors endogenously become more elastic: they become more willing to absorb shocks, which expands market absorption capacity and lowers the multiplier.

7 Conclusion

Large demand shocks expand rather than exhaust the capacity of stock markets to absorb quantities. We show that price multipliers for individual stocks, which capture the compensation investors require to absorb shocks, are smaller for larger contemporaneous and cumulative past shocks. This pattern is robust across three uninformed demand shocks from previous work: Russell index-reconstitution induced changes in benchmarking intensity, fund flow-induced trading, and order flow imbalance. Consistent with these results, we find that institutional investors become more price elastic as price dislocations grow.

Thus, investors endogenously become more elastic — more willing to absorb shocks — when larger shocks create larger price dislocations and larger profit opportunities. As a result, market absorption capacity expands and multipliers fall. We provide two examples of mechanisms that give rise to this behavior: fixed adjustment costs and endogenous inattention.

Our findings suggest that price multipliers for the large shocks of interest in many finance questions — such as government asset purchases, green investing, and share buybacks and issuances — are likely smaller than multiplier estimates based on small shocks.

Moreover, our findings provide guidance for future work that seeks to model investors' asset demand. While previous work on asset demand systems often assumes fixed elasticities, our results suggest that incorporating the endogenous behavior of these elasticities can allow for a more realistic characterization of demand and price dynamics. Thus, enriching models with endogenous elasticities can provide more accurate answers to important questions regarding the impact of shifts in quantities on financial markets.

References

Alfonsi, A., A. Fruth, and A. Schied. 2010. Optimal execution strategies in limit order books with general shape functions. *Quantitative finance* 10:143–157.

Amihud, Y., and H. Mendelson. 1980. Dealership market: Market-making with inventory. *Journal of financial economics* 8:31–53.

An, Y., Y. Su, and C. Wang. 2025. Quantity, risk, and return. *Available at SSRN 4098609* .

Appel, I., T. A. Gormley, and D. B. Keim. 2021. Identification using Russell 1000/2000 index assignments: A discussion of methodologies. *Critical Finance Review, Forthcoming* .

Appel, I. R., T. A. Gormley, and D. B. Keim. 2019. Standing on the shoulders of giants: The effect of passive investors on activism. *The Review of Financial Studies* 32:2720–2774.

Ben-David, I., F. Franzoni, and R. Moussawi. 2019. An improved method to predict assignment of stocks into Russell indexes. Tech. rep., National Bureau of Economic Research.

Bouchaud, J.-P., J. Bonart, J. Donier, and M. Gould. 2018. *Trades, quotes and prices: financial markets under the microscope.* Cambridge University Press.

Bretscher, L., L. Schmid, I. Sen, and V. Sharma. 2025. Institutional corporate bond pricing. *The Review of Financial Studies* p. hhaf067.

Brunnermeier, M. K., and L. H. Pedersen. 2009. Market liquidity and funding liquidity. *The review of financial studies* 22:2201–2238.

Chacko, G. C., J. W. Jurek, and E. Stafford. 2008. The price of immediacy. *The Journal of Finance* 63:1253–1290.

Chang, Y.-C., H. Hong, and I. Liskovich. 2015. Regression discontinuity and the price effects of stock market indexing. *The Review of Financial Studies* 28:212–246.

Chaudhary, M., Z. Fu, and H. Zhou. 2024. Anatomy of the Treasury Market: Who Moves Yields? *Available at SSRN* .

Chaudhry, A. 2023. The Impact of Prices on Analyst Cash Flow Expectations. *Working paper* .

Chen, A. Y., and T. Zimmermann. 2022. Open Source Cross-Sectional Asset Pricing. *Critical Finance Review* 27:207–264.

Chen, H. 2024. Diversification Driven Demand for Large Stock. *Available at SSRN 4288953* .

Constantinides, G. M. 1986. Capital market equilibrium with transaction costs. *Journal of political Economy* 94:842–862.

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

Crane, A. D., S. Michenaud, and J. P. Weston. 2016. The effect of institutional ownership on payout policy: Evidence from index thresholds. *The Review of Financial Studies* 29:1377–1408.

Cremers, K. M., and A. Petajisto. 2009. How active is your fund manager? A new measure that predicts performance. *The review of financial studies* 22:3329–3365.

Davis, C., M. Kargar, and J. Li. 2025. Why do portfolio choice models predict inelastic demand? *Journal of Financial Economics* 172:104096.

Davis, S. J., and J. Haltiwanger. 1992. Gross job creation, gross job destruction, and employment reallocation. *The Quarterly Journal of Economics* 107:819–863.

Diamond, D. W., and R. E. Verrecchia. 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of financial economics* 18:277–311.

Donier, J., J. Bonart, I. Mastromatteo, and J.-P. Bouchaud. 2015. A fully consistent, minimal model for non-linear market impact. *Quantitative finance* 15:1109–1121.

Duffie, D., N. Garleanu, and L. H. Pedersen. 2002. Securities lending, shorting, and pricing. *Journal of Financial Economics* 66:307–339.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of political economy* 81:607–636.

Frazzini, A., R. Israel, and T. J. Moskowitz. 2018. Trading Costs. *Working paper* .

Gabaix, X. 2019. Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1*, vol. 2, pp. 261–343. Elsevier.

Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley. 2006. Institutional investors and stock market volatility. *The Quarterly Journal of Economics* 121:461–504.

Gabaix, X., and R. S. Koijen. 2022. In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis. *Working paper* .

Gabaix, X., and R. S. Koijen. 2024. Granular instrumental variables. *Journal of Political Economy* 132:2274–2303.

Gatheral, J. 2010. No-dynamic-arbitrage and market impact. *Quantitative finance* 10:749–759.

Glossner, S. 2019. Russell index reconstitutions, institutional investors, and corporate social responsibility. *Available at SSRN 3180776* .

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift. 2020. Bartik instruments: What, when, why, and how. *American Economic Review* 110:2586–2624.

Gromb, D., and D. Vayanos. 2002. Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of financial Economics* 66:361–407.

Grossman, S. J., and G. Laroque. 1987. Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods.

Grossman, S. J., and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *The American economic review* 70:393–408.

Haddad, V., Z. He, P. Huebner, P. Kondor, and E. Loualiche. 2025. Causal inference for asset pricing. *Available at SSRN 5187305* .

Haddad, V., P. Huebner, and E. Loualiche. 2024. How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing. *Working paper* .

Han, X., N. L. Roussanov, and H. Ruan. 2021. Mutual fund risk shifting and risk anomalies. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper* .

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society* pp. 1029–1054.

Harris, L., and E. Gurel. 1986. Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures. *the Journal of Finance* 41:815–829.

Hartzmark, S. M., and D. H. Solomon. 2022. Predictable price pressure. Tech. rep., National Bureau of Economic Research.

He, Z., P. Kondor, and J. S. Li. 2025. Demand Elasticity in Dynamic Asset Pricing. *Available at SSRN 5297190* .

Hong, H., and J. C. Stein. 2003. Differences of opinion, short-sales constraints, and market crashes. *The Review of Financial Studies* 16:487–525.

Huang, S., Y. Song, and H. Xiang. 2024. Noise trading and asset pricing factors. *Management Science* .

Huberman, G., and W. Stanzl. 2004. Price manipulation and quasi-arbitrage. *Econometrica* 72:1247–1275.

Huebner, P. 2023. The making of momentum: A demand-system perspective. In *Proceedings of the EUROFIDAI-ESSEC Paris December Finance Meeting*.

Jansen, K. A. 2025. Long-term investors, demand shifts, and yields. *The Review of Financial Studies* 38:114–157.

Jansen, K. A., W. Li, and L. Schmid. 2024. Granular treasury demand with arbitrageurs. Tech. rep., National Bureau of Economic Research.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp. 2016. A rational theory of mutual funds' attention allocation. *Econometrica* 84:571–626.

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

Koijen, R. S., R. J. Richmond, and M. Yogo. 2024. Which investors matter for equity valuations and expected returns? *Review of Economic Studies* 91:2387–2424.

Koijen, R. S., and M. Yogo. 2019. A demand system approach to asset pricing. *Journal of Political Economy* 127:1475–1515.

Koijen, R. S., and M. Yogo. 2020. Exchange rates and asset prices in a global demand system. Tech. rep., National Bureau of Economic Research.

Kyle, A. S. 1989. Informed speculation with imperfect competition. *The Review of Economic Studies* 56:317–355.

Kyle, A. S., and W. Xiong. 2001. Contagion as a wealth effect. *The Journal of Finance* 56:1401–1440.

Lee, C. M., and M. J. Ready. 1991. Inferring trade direction from intraday data. *The Journal of Finance* 46:733–746.

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

Li, J., and Z. Lin. 2025. Price Multipliers are Larger for Less Diversifiable Order Flows. *Working paper* .

Liu, H. 2004. Optimal consumption and investment with transaction costs and multiple risky assets. *The Journal of Finance* 59:289–338.

Lo, A. W., H. Mamaysky, and J. Wang. 2004. Asset prices and trading volume under fixed transactions costs. *Journal of Political Economy* 112:1054–1090.

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

Pavlova, A., and T. Sikorskaya. 2023. Benchmarking intensity. *The Review of Financial Studies* 36:859–903.

Petajisto, A. 2009. Why do demand curves for stocks slope down? *Journal of Financial and Quantitative Analysis* 44:1013–1044.

Pollet, J. M., and M. Wilson. 2008. How does size affect mutual fund behavior? *The Journal of finance* 63:2941–2969.

Scarf, H. 1960. The optimality of (S, s) policies in the dynamic inventory problem. *Mathematical Methods in the Social Sciences* pp. 196–202.

Schmickler, S., and P. Tremacoldi-Rossi. 2022. Spillover effects of payouts on asset prices and real investment. *Available at SSRN 4287300* .

Shleifer, A. 1986. Do demand curves for stocks slope down? *The Journal of Finance* 41:579–590.

Sims, C. A. 2003. Implications of rational inattention. *Journal of monetary Economics* 50:665–690.

Tamoni, A., S. Sokolinski, and Y. Li. 2024. Which Investors Drive Anomaly Returns and How? *Available at SSRN 4242745* .

Tóth, B., Y. Lemperiere, C. Deremble, J. De Lataillade, J. Kockelkoren, and J.-P. Bouchaud. 2011. Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X* 1:021006.

van Binsbergen, J. H., B. David, and C. C. Opp. 2025. How (not) to identify demand elasticities in dynamic asset markets. Tech. rep., National Bureau of Economic Research.

Van der Beck, P. 2025. Short versus Long-run Demand Elasticities in Asset Pricing. *Working paper* .

Van Nieuwerburgh, S., and L. Veldkamp. 2009. Information immobility and the home bias puzzle. *The Journal of Finance* 64:1187–1215.

Van Nieuwerburgh, S., and L. Veldkamp. 2010. Information acquisition and under-diversification. *The Review of Economic Studies* 77:779–805.

Wei, W., and A. Young. 2021. Selection bias or treatment effect? A re-examination of Russell 1000/2000 Index reconstitution. *Critical Finance Review* 1000.

Zhou, Y., and Y. Chen. 2025. Deflated HeteroPCA: Overcoming the curse of ill-conditioning in heteroskedastic PCA. *The Annals of Statistics* 53:91–116.

APPENDIX

A Supplemental Results for Demand Shock Measurement

This section provides supplemental results and discussions about the three demand shocks introduced in Section 3.2.

A.1 Description of Russell Banding Methodology Starting in 2007

This appendix describes the methodology used to calculate the market cap rank cutoffs for the annual Russell index reconstitutions.

Prior to 2007, Russell reconstitution rules were simple. Firms with market capitalizations on the May rank date that fell between ranks 1 and 1000 were assigned to the Russell 1000, and those with market caps ranked between 1001 and 3000 were assigned to the Russell 2000.

To reduce turnover, since 2007 FTSE Russell has used a "banding policy" under which there are two separate cutoffs for stocks in the Russell 1000 and 2000 in the previous year, both of which are based on a mechanical function of the firm size distribution in the year. Under this policy:

- Stocks in the Russell 2000 in the previous year are assigned to the Russell 1000 if their rank date market cap ranks fall between 1 and $1000 - c_1$.
- Stocks in the Russell 1000 in the previous year are assigned to the Russell 2000 if their rank date market cap ranks fall between $1000 + c_2$ and 3000.

To calculate c_1 and c_2 Russell first computes the cumulative market cap of the largest 1000 stocks (i.e., those with ranks 1 through 1000). Let $C(N)$ represent the cumulative market cap of the largest N stocks. c_1 is calculated such that $C(1000 - c_1) = 0.95 \cdot C(1000)$. c_2 is calculated such that $C(1000 + c_2) = 1.05 \cdot C(1000)$. That is, the band of stocks between ranks $1000 - c_1$ and $1000 + c_2$ constitutes a 5% band around the cumulative market cap of the largest 1000 stocks.

Thus, even after the introduction of the banding policy, assignment to the Russell 1000 or 2000 is still based on a mechanical rule. After the introduction of the banding policy, this mechanical rule changes each year with the distribution of firm sizes.

A.2 FIT Measurement

This section describes how we construct the flow-induced trading (FIT) demand shock. In particular, we modify the construction in Lou (2012) to yield a cleaner FIT shock in order to help alleviate measurement error concerns. We do this modification to guard against our main result — that multipliers are smaller when the magnitude of FIT is large — arises from attenuation bias. Therefore, we would like to alleviate the concern of overestimating large FIT magnitudes in the tails, which would lead to stronger attenuation bias for larger shocks in the price multiplier regressions. All our results are not sensitive to whether we apply this cleaning procedure.

Where would mismeasurement come from? Under the simplest specification where we assume that fund trades respond one-to-one to fund flows, the flow-induced trading of fund i in stock n, in number of shares, is given by

$$\text{SharesHeld}_{i,n,t-1} \times f_{i,t}. \tag{A.1}$$

where $\text{SharesHeld}_{i,n,t-1}$ is the lagged number of shares held by fund i in stock n and $f_{i,t}$ is the flow in fund i, expressed as a fraction of lagged assets under management. Equation (A.1) shows that there are two possible sources of mismeasurement, which we tackle in turn.

1. **Heterogeneous trading response to fund flows.** Fund trades may respond less than one-to-one to fund flows, and this may be especially true when flow magnitudes are large or when pre-existing positions are large, both of which can lead us to overestimate large FIT values.[1]

   We address this concern by estimating heterogeneous flow-to-trade responses via a panel regression of trades on dummy variables:

$$\text{Trade}_{i,n,t} = \sum_b \sum_f \beta_{b,f} \cdot \text{I}_{w_{i,n,t-1} \in \text{ bin } b} \times \text{I}_{f_{i,t} \in \text{ bin } f} + \epsilon_{i,n,t} \tag{A.2}$$

   where $\text{Trade}_{i,n,t}$ is defined as $\frac{\text{SharesHeld}_{i,n,t}}{\text{SharesHeld}_{i,n,t-1}} - 1$ and $w_{i,n,t-1}$ is the lagged portfolio weight of stock n for fund i. We obtain quarterly fund flows of domestic equity funds from

[1] For instance, when a fund faces large inflows, it may use a higher fraction of the new capital to purchase stocks it did not previously hold, which reduces the need to purchase stocks in the existing holdings (e.g., Pollet and Wilson, 2008; Lou, 2012). Further, for diversification purposes, it may reduce its purchase if a stock already occupies a large part of the portfolio (Chen, 2024).

CRSP and merge it with quarterly trades in 13F filings.

To study heterogeneous responses, we sort the full sample by fund flows into $f = -20, ..., 0, ..., 20$ bins, with the first (last) 20 bins covering the flows below -1% (above 1%) realizations, and bin 0 is defined by $f_{i,t} \in [-1\%, 1\%]$. Panel (a) of Figure A.1 plots the average flow by bins and shows that they cover a large range from approximately -30% to +100%.

Similarly, for each fund in each period, we sort its pre-existing stock holdings by the portfolio weight $w_{i,n,t-1}$ into $b = 1, ..., 20$ bins. To reduce the impact of outliers due to "dividing by a small number" in the dependent variable, we estimate (A.2) via a weighted regression with weights equal to $w_{i,n,t-1}$.[2]

Panel (b) of Figure A.1 plots the point estimates of $\beta_{b,f}$ for bins $b = 1, 10$, and 20. The results indeed indicate that when facing inflows, funds tend to trade less than one-to-one in stocks that they already have large holdings in, but the heterogeneity is limited in magnitude. Even for the top bin (blue line), the flow-to-trade pass-through is over 80%. The flow-to-trade pass-through is closer to one-to-one for outflows, a finding that is consistent with Lou (2012).

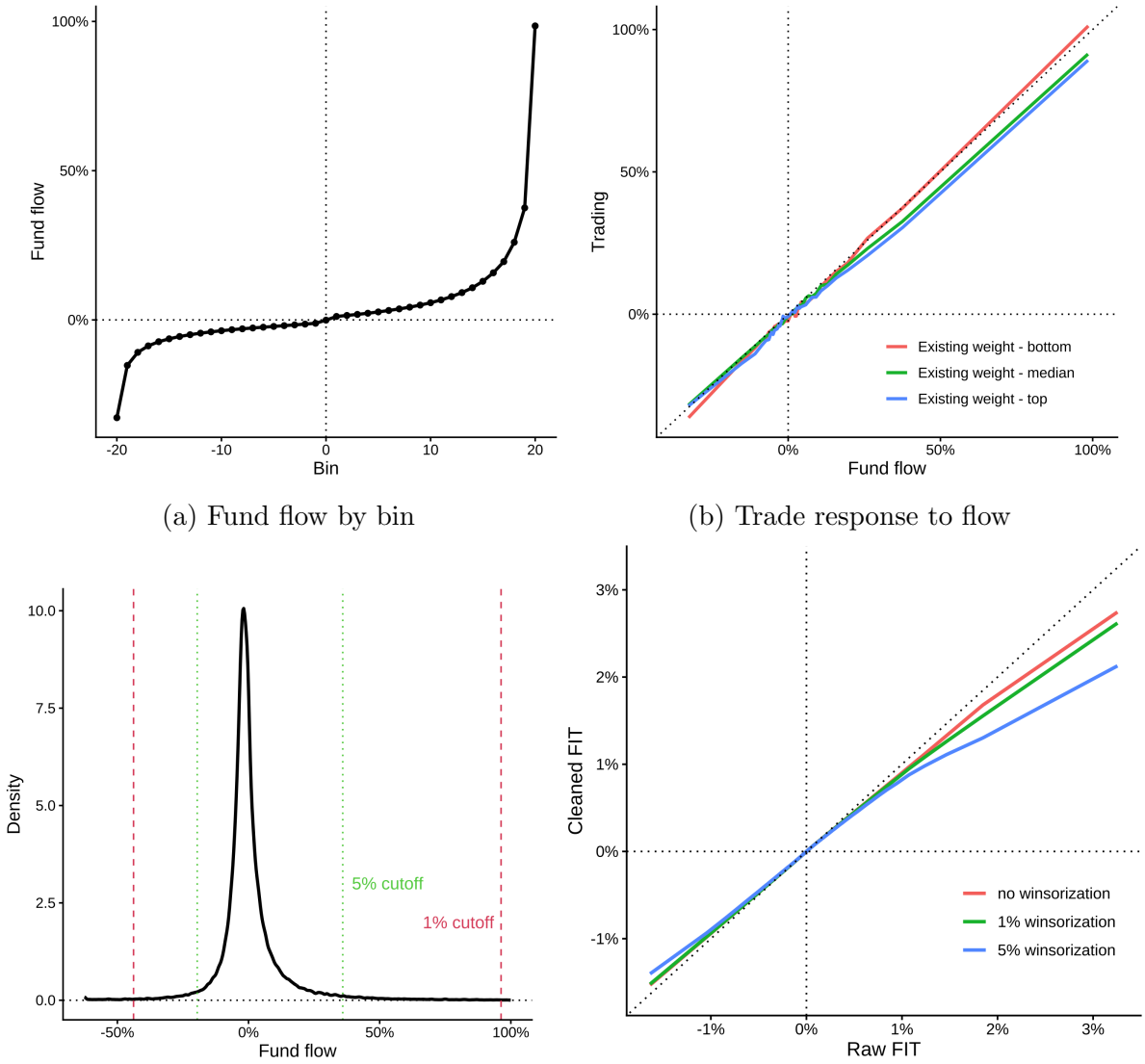
To account for the heterogeneity of flow-to-trade responses, we compute FIT as:

$$\text{FIT}_{n,t} = \sum_{\text{fund } i} \frac{\text{SharesHeld}_{i,n,t-1}}{\text{Share Outstanding}_{n,t-1}} \cdot \underbrace{\beta(w_{i,n,t-1}, f_{i,t})}_{\text{estimated in (A.2)}}$$

where the heterogeneous responses $\beta(w_{i,n,t-1}, f_{i,t})$ are the point estimates from regression (A.2).

2. **Winsorize fund flows.** Another possible concern is that the extreme fund flows themselves may be misestimated. To alleviate this concern, we recompute FIT after winsorizing a total of 1% (or 5%) of extreme observations in fund flows, with equal winsorization on each side. The winsorization thresholds are illustrated in Panel (c) of Figure A.1. The black line plots the density of fund flows, and the colored vertical dashed lines represent the cutoffs. By removing extreme values of fund flows, we ensure that our FIT measure is not driven by mismeasured large flow magnitudes.

[2]To control for the fact that, even without flows, portfolio weights tend to mean-revert — that is, the largest (smallest) positions tend to be reduced (increased) subsequently — we estimate a first-stage regression where we regress the dependent variable on indicators of existing portfolio weight bin b. We use the resulting residuals as the dependent variable in regression (A.2).



(a) Fund flow by bin

(b) Trade response to flow

(c) Fund flow winsorization

(d) Comparing FIT measures

Figure A.1. Cleaning the flow-induced trading measure

In Panel (a), we sort realized fund flows into bins in each cross-section and plot the average flows by bins. Panel (b) plots the average fund trade as a function of fund flows by the size of the pre-existing portfolio weights; the dashed diagonal line is the 45 degree line. Panel (c) plots the kernel density of fund flows, and the vertical dashed lines represent the various winsorization cutoffs. Panel (d) plots the cleaned versions of FIT against the "raw FIT" which does not winsorize fund flows and assumes that trades respond one-to-one to flows.

Panel (d) of Figure A.1 shows the effect of applying these cleaning and winsorization steps. We sort the sample into 100 bins by "raw FIT," which does not winsorize flows and assumes a one-to-one flow-to-trade response as in (1), and we plot it on the horizontal axis. On the vertical axis, the red line plots the average FIT after taking into account heterogeneous flow-to-trade responses. The remaining lines plot the results after further applying fund flow winsorization. The plot suggests that applying these cleaning steps serves to gradually dampen the large values of FIT. In the main paper, we use the version represented by the green line, but our main results are robust to all of these variants of the FIT measure.

B Additional Empirical Results for Section 4

B.1 Differential Correlation with Unobserved Demand

In this section, we test the alternative hypothesis that our results are driven by a differential degree of correlation between our demand shocks and other investors' trading. Our demand shocks do not capture all trading; unobserved trading correlated with our shocks may constitute an omitted variable. The specific concern we aim to address is the possibility that, when the magnitude of our measured demand shock is smaller (larger), there is more (less) correlated unobserved trading in the market. This situation could spuriously give rise to our results even if the true price multiplier is constant.

BMI. Pavlova and Sikorskaya (2023) show that changes in BMI — driven by exogenous Russell reconstitutions — induce changes in institutional ownership of stocks. In the context of our paper, one may be concerned that when BMI changes have larger magnitudes, their "pass-through rate" to institutional ownership changes is lower, which can explain our finding even if the true price multiplier is constant.

In Table B.1, we test this concern using Fama-MacBeth regressions which estimate the effect of ΔBMI on quarterly 13F institutional ownership changes of stocks (as a fraction of lagged shares outstanding). Column (1) controls for the same set of controls as in Pavlova and Sikorskaya (2023), and the subsequent two columns progressively add the predictor and liquidity controls in Table 2.

This result can be compared to Table 3 of Pavlova and Sikorskaya (2023) (column 3, panel B). We verify their finding that BMI has a clear effect on institutional holdings. Our estimated coefficient is slightly higher than theirs: recall that we have multiplied by the original BMI measure by 0.2. Therefore, exactly replicating their result implies a pass-through rate

coefficient of one. In our analysis, we find slightly higher coefficients, implying an even stronger pass-through rate of BMI to institutional ownership. Yet, crucially, we do not find evidence that the BMI pass-through rate declines with shock size. For instance, throughout the different regression specifications, the coefficient for BMI when its magnitude is lower than one standard deviation is not statistically or economically different from that when its magnitude is above two standard deviations.

Panel A: BMI pass-through regression			
	Dependent variable: $\Delta IO_{n,t}$		
	(1)	(2)	(3)
$M_{\lvert d_{n,t}\rvert<\sigma}$	1.50***	1.36***	1.19**
	(0.47)	(0.47)	(0.47)
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	1.77***	1.63***	1.48***
	(0.47)	(0.47)	(0.47)
$M_{\lvert d_{n,t}\rvert>2\sigma}$	1.42***	1.34***	1.26***
	(0.47)	(0.47)	(0.47)
Pavlova and Sikorskaya (2023) controls	Y	Y	Y
Predictor controls	N	Y	Y
Liquidity controls	N	N	Y
Obs	9,565	9,565	9,565
R^2	0.067	0.113	0.149
Panel B: Coefficient differences			
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]} - M_{\lvert d_{n,t}\rvert>2\sigma}$	0.36	0.30	0.23
	(0.51)	(0.49)	(0.48)
$M_{\lvert d_{n,t}\rvert<\sigma} - M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	−0.27	−0.27	−0.29
	(0.60)	(0.63)	(0.61)
$M_{\lvert d_{n,t}\rvert<\sigma} - M_{\lvert d_{n,t}\rvert>2\sigma}$	0.09	0.03	−0.06
	(0.66)	(0.64)	(0.64)

Table B.1. Pass-through Rate of ΔBMI to institutional ownership changes

We use Fama-MacBeth regressions to estimate the effect of ΔBMI on contemporaneous quarterly change in 13F instititional ownership of stocks. The columns differ by the controls utilized. Panel A reports regression coefficients. Panel B reports pairwise differences in price multipliers with the standard error computed using the Delta method. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

FIT. One may be concerned that other trading imbalances may be correlated with FIT. It is plausible that the same common factors that drive flows to different mutual funds also drive trading among non-mutual fund investors. For example, in periods where technology sector funds receive high inflows from retail investors, perhaps those retail investors are conducting similar allocations in their personal portfolios. Thus, removing the common factors that drive flows to different mutual funds and only using idiosyncratic fund flows to construct FIT would remove this correlation with non-mutual-fund trades (under the assumption that the same factors drive non-mutual-fund trades as well).

To examine if our results are robust to this concern, we conduct a principal component analysis (PCA) on the fund-by-quarter panel of fund flows. We indeed observe sizeable common variation: the first ten PCs can explain almost half of cross-sectional fund flow variation. Because the flow residual panel is unbalanced, we conduct PCA only on the non-missing elements of the panel using alternating least squares. We then remove the top ten PCs and use the fund flow residuals to compute stock-level FIT, following equation (1).

Figure B.2 visualizes the differences in the stock-level FIT after removing fund flow PCs. The original FIT is plotted on the horizontal axis and the FIT after removing PCs is plotted on the vertical axis. Removing PCs indeed reduced the variance of FIT, and the figure suggests that the effect starts to stablize after removing 10 PCs.

We then re-estimate price multipliers using the PC-removed FIT; the regression specification and the set of controls is the same as column (6) in Table 2. Results are reported in Table B.2. For comparison, the results with the original FIT is reported in column (1), and the subsequent columns report results based on FIT versions after removing more PCs. Throughout all these specifications, we continue to see clear evidence that multipliers are smaller when FIT magnitude is larger.

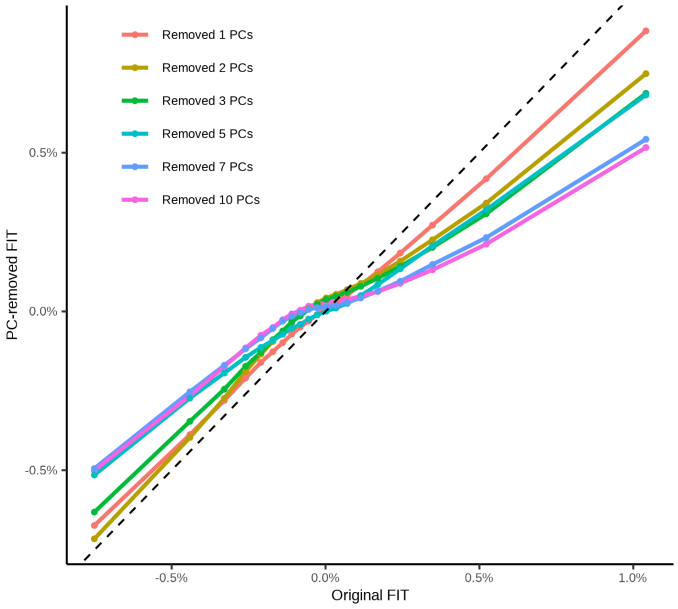


Figure B.2. FIT: removing common components in flows

As described in Appendix B.1, we remove up to ten principal components in fund flows before computing FIT. This figure visualizes the effect on the FIT mesure. The horizonal axis plots the original FIT and the vertical axis plots the FIT after removing flow PCs.

B.2 Robustness Checks for Static Results

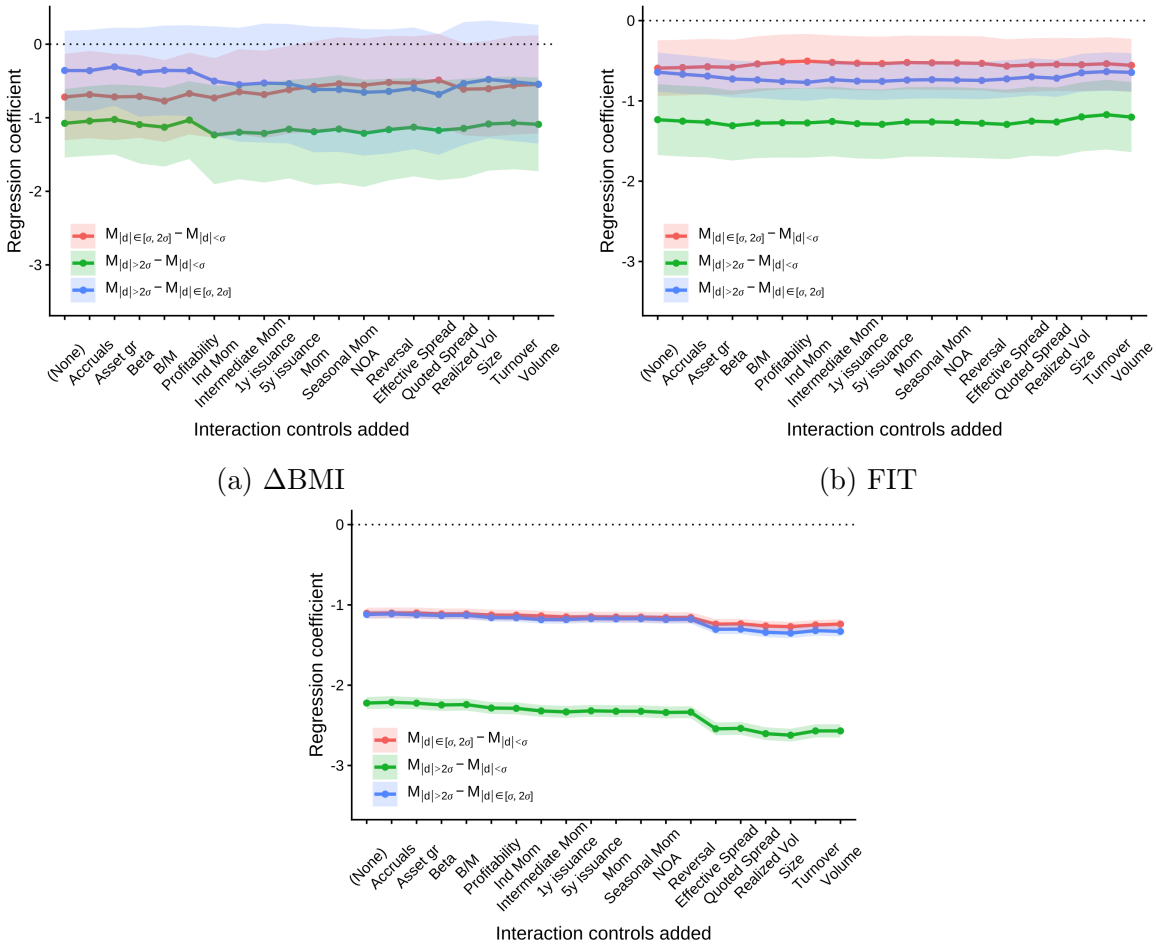
Panel regressions. Table B.3 uses panel regressions with time fixed effects to estimate regression (2), and we cluster standard errors by time and stock. The results are consistent with the Fama-MacBeth results in Table 2, indicating that price multipliers are smaller when demand shocks are larger in magnitude.

Panel A: price impact regressions							
	Dependent variable: stock return $r_{n,t}$						
PCs removed	0	1	2	3	5	7	10
	(1)	(2)	(3)	(4)	(5)		
$M_{\lvert d_{n,t}\rvert<\sigma}$	3.92***	3.25***	3.57***	2.78***	2.25***	2.87***	3.26***
	(0.52)	(0.46)	(0.46)	(0.46)	(0.44)	(0.42)	(0.39)
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	3.33***	2.91***	2.66***	2.31***	2.13***	2.07***	2.09***
	(0.30)	(0.29)	(0.30)	(0.32)	(0.29)	(0.27)	(0.26)
$M_{\lvert d_{n,t}\rvert>2\sigma}$	2.69***	2.25***	2.05***	1.83***	1.61***	1.72***	1.61***
	(0.26)	(0.23)	(0.20)	(0.22)	(0.22)	(0.20)	(0.19)
Predictor controls	Y	Y	Y	Y	Y	Y	Y
Liquidity controls	Y	Y	Y	Y	Y	Y	Y
Obs	561,404	545,103	545,103	545,103	545,103	545,103	545,103
R^2	0.080	0.080	0.080	0.079	0.079	0.078	0.078
Marginal $R^2(d_{n,t})$	0.005	0.004	0.004	0.004	0.003	0.003	0.003
Panel B: Coefficient differences							
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]} - M_{\lvert d_{n,t}\rvert<\sigma}$	−0.59*	−0.34	−0.92***	−0.47	−0.11	−0.79**	−1.17***
	(0.35)	(0.31)	(0.32)	(0.31)	(0.31)	(0.34)	(0.32)
$M_{\lvert d_{n,t}\rvert>2\sigma} - M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	−0.64***	−0.66***	−0.61***	−0.48**	−0.53**	−0.35	−0.48**
	(0.24)	(0.23)	(0.23)	(0.22)	(0.25)	(0.24)	(0.21)
$M_{\lvert d_{n,t}\rvert>2\sigma} - M_{\lvert d_{n,t}\rvert<\sigma}$	−1.23***	−1.01***	−1.53***	−0.95**	−0.64*	−1.15***	−1.65***
	(0.44)	(0.39)	(0.40)	(0.38)	(0.39)	(0.37)	(0.36)

Table B.2. FIT multiplier by contemporaneous demand size: removing PCs

This table examples the robustness of FIT-based inference. As described in Appendix B.1, we progressively remove 1 to 10 principal components (PCs) in fund flows and use the residual flows to compute FIT. We then re-estimate the Fama-MacBeth regression of column (6) in Table 2 which is reproduced as column (1) in this table. The subsequent columns in this table differ by the number of PCs removed. Panel A reports regression coefficients. Panel B reports pairwise differences in price multipliers with the standard error computed using the Delta method. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Interaction with stock characteristics. In Section 4.2, we estimate regression (6) to investigate whether controlling for demand shock interactions with stock characteristics influence our findings. To visualize the impact of progressively adding each specific characteristic, Figure B.3 displays the estimated multiplier differences (e.g. $M_{|d_{n,t}|>2\sigma} - M_{|d_{n,t}|\in[\sigma,2\sigma]}$) from these regressions. Each point represents the estimates from regression (6) when adding one additional characteristic interaction. Specifically, the left-most points represent coefficient estimates without any interacted controls, and the second left-most points represent estimates after controlling for $d_{n,t} \times \text{Accruals}_{n,t}$. The right-most points represent coefficient estimates from the regression including all characteristics interacted with demand. The shaded areas represent one standard error bands. We find that the estimates are relatively stable throughout the process of adding controls.



(a) ΔBMI

(b) FIT

(c) OFI

Figure B.3. Multiplier differences: adding demand-interacted controls

Panels (a), (b), and (c) plot the estimated pairwise differences of price multipliers (e.g. $M_{|d_{n,t}|>2\sigma} - M_{|d_{n,t}|\in[\sigma,2\sigma]}$) in regression (6) for the ΔBMI, FIT, and OFI demand shocks, respectively. The left-most points represent estimates when no interactions between the demand shock and any stock characteristics are included. Each subsequent point represents the estimates from regression (6) when adding the demand interaction with characteristic labelled on the x-axis (i.e., the right-most points represent coefficient estimates from the regression including all liquidity and stock characteristics). The shaded areas represent one standard error bands.

Panel A: price impact regressions									
	Dependent variable: stock return $r_{n,t}$								
	BMI			FIT			OFI		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$M_{\lvert d_{n,t}\rvert<\sigma}$	2.61***	2.46***	2.37***	6.95***	6.91***	6.97***	2.89***	3.04***	3.41***
	(0.66)	(0.64)	(0.62)	(1.26)	(1.23)	(1.25)	(0.12)	(0.13)	(0.16)
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	1.31*	1.18*	1.14*	4.65***	4.64***	4.64***	1.86***	1.94***	2.17***
	(0.69)	(0.63)	(0.63)	(0.71)	(0.70)	(0.70)	(0.11)	(0.11)	(0.12)
$M_{\lvert d_{n,t}\rvert>2\sigma}$	0.77**	0.71**	0.67**	3.26***	3.25***	3.25***	0.79***	0.83***	0.94***
	(0.31)	(0.31)	(0.33)	(0.41)	(0.40)	(0.39)	(0.07)	(0.07)	(0.07)
Predictor controls	N	Y	Y	N	Y	Y	N	Y	Y
Liquidity controls	N	N	Y	N	N	Y	N	N	Y
Obs	9,910	9,910	9,910	561,404	561,404	561,404	527,744	527,744	527,744
R^2	0.140	0.155	0.157	0.153	0.157	0.157	0.189	0.196	0.202
Marginal $R^2(d_{n,t})$	0.006	0.005	0.004	0.004	0.004	0.004	0.040	0.042	0.047
Panel B: Coefficient differences									
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]} - M_{\lvert d_{n,t}\rvert<\sigma}$	−1.29*	−1.28*	−1.22*	−2.30***	−2.27***	−2.33***	−1.03***	−1.10***	−1.24***
	(0.73)	(0.73)	(0.71)	(0.64)	(0.62)	(0.64)	(0.06)	(0.07)	(0.08)
$M_{\lvert d_{n,t}\rvert>2\sigma} - M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	−0.54	−0.48	−0.47	−1.39***	−1.39***	−1.40***	−1.07***	−1.11***	−1.23***
	(0.65)	(0.60)	(0.61)	(0.38)	(0.38)	(0.39)	(0.06)	(0.06)	(0.06)
$M_{\lvert d_{n,t}\rvert>2\sigma} - M_{\lvert d_{n,t}\rvert<\sigma}$	−1.84***	−1.76***	−1.69***	−3.69***	−3.67***	−3.72***	−2.10***	−2.21***	−2.46***
	(0.71)	(0.66)	(0.64)	(0.93)	(0.91)	(0.94)	(0.08)	(0.09)	(0.11)

Table B.3. Price multiplier by contemporaneous demand size: panel regression

This table is similar to Table 2 except that we estimate regression (2) using panel regressions. Standard errors are clustered by time and stock. Panel A reports regression coefficients. Panel B reports pairwise differences in price multipliers with the standard error computed using the Delta method. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Standardizing demand shocks by stock-specific demand volatility. Following the discussion in Section 4.2, we standardize FIT and OFI using stock-specific demand volatilities. We do not conduct the same exercise with BMI because it only realizes once a year, and so we cannot estimate demand volatility with precision. Specifically, for each demand measure $d_{n,t}$, we use the previous h =4, 8, or 12 quarters to estimate stock-specific demand volatilities, which we normalize demand shocks by. To make the resulting demand variable have equal variability as the original demand, we also multiply demand shocks by the full-sample demand standard deviation. This ensures that the resulting regression coefficients can be quantitatively compared to the results before standardizing demand.

We then use the standardized demand shocks $d_{n,t}^{\text{std}}$ instead of the raw demand shocks $d_{n,t}$ in regression (2), and we report the results in Table B.4. Columns (1) through (4) study FIT and columns (5) through (8) study OFI. For comparison purposes, columns (1) and (4) report the original results from Table 3 with full controls. The other columns report regression results based on standardized demand and are qualitatively similar.

Panel A: price impact regressions								
	Dependent variable: stock return $r_{n,t}$							
	FIT				OFI			
Lookback window	N/A	4	8	12	N/A	4	8	12
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$M_{\lvert d^{\text{std}}_{n,t}\rvert<\sigma}$	3.76***	4.87***	4.06***	4.07***	3.15***	3.08***	2.81***	2.73***
	(0.52)	(0.48)	(0.46)	(0.42)	(0.10)	(0.13)	(0.12)	(0.12)
$M_{\lvert d^{\text{std}}_{n,t}\rvert\in[\sigma,2\sigma]}$	3.21***	3.31***	3.17***	2.89***	1.91***	1.70***	1.76***	1.80***
	(0.33)	(0.30)	(0.31)	(0.29)	(0.07)	(0.07)	(0.08)	(0.08)
$M_{\lvert d^{\text{std}}_{n,t}\rvert>2\sigma}$	2.56***	1.28***	1.74***	2.13***	0.58***	0.53***	0.65***	0.71***
	(0.24)	(0.17)	(0.19)	(0.21)	(0.08)	(0.05)	(0.06)	(0.07)
Predictor controls	Y	Y	Y	Y	Y	Y	Y	Y
Liquidity controls	Y	Y	Y	Y	Y	Y	Y	Y
Interacted: predictors	Y	Y	Y	Y	Y	Y	Y	Y
Interacted: liquidity	Y	Y	Y	Y	Y	Y	Y	Y
Obs	561,404	491,638	436,750	392,262	527,744	425,234	352,302	295,844
R^2	0.087	0.091	0.095	0.096	0.155	0.153	0.162	0.168
Panel B: Coefficient differences								
$M_{\lvert d^{\text{std}}_{n,t}\rvert\in[\sigma,2\sigma]}$ - $M_{\lvert d^{\text{std}}_{n,t}\rvert<\sigma}$	−0.56*	−1.56***	−0.89**	−1.19***	−1.24***	−1.38***	−1.05***	−0.93***
	(0.33)	(0.44)	(0.36)	(0.29)	(0.06)	(0.07)	(0.06)	(0.06)
$M_{\lvert d^{\text{std}}_{n,t}\rvert>2\sigma}$ - $M_{\lvert d^{\text{std}}_{n,t}\rvert\in[\sigma,2\sigma]}$	−0.65***	−2.03***	−1.42***	−0.76***	−1.33***	−1.17***	−1.10***	−1.09***
	(0.24)	(0.28)	(0.28)	(0.25)	(0.06)	(0.05)	(0.05)	(0.05)
$M_{\lvert d^{\text{std}}_{n,t}\rvert>2\sigma}$ - $M_{\lvert d^{\text{std}}_{n,t}\rvert<\sigma}$	−1.20***	−3.59***	−2.31***	−1.94***	−2.57***	−2.54***	−2.15***	−2.02***
	(0.44)	(0.46)	(0.43)	(0.39)	(0.08)	(0.10)	(0.08)	(0.08)

Table B.4. Multiplier by contemporaneous demand sizes: standardized demand

This table reports regression results for FIT and OFI and is similar to Table 3, except that it uses standardized demand ($d^{\text{std}}_{n,t}$) instead of raw demand ($d_{n,t}$). Standardized demand is computed as $d^{\text{std}}_{n,t} = \frac{d_{n,t}}{\hat{\sigma}_{(h)}(d_{n,t})} \times \bar{\sigma}$, where $\hat{\sigma}_{(h)}(d_{n,t}) = \sqrt{\frac{1}{h-1}\sum_{l=1}^{h}(d_{n,t-l} - \bar{d}_{n,t-1})^2}$ with $h = 4, 8,$ or 12 as the lookback window, and where we multiply the standardized demand with the full-sample volatility ($\bar{\sigma}$) so that the resulting standardized demand has the same amount of variation as the original one. For comparison purposes, columns (1) and (5) reproduce columns (6) and (9) of Table 3 where the demand is not standardized. In the other columns, we report results using standardized demand with different lookback windows h. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Anticipation of demand shocks. As discussed in Section 4.2, one might be concerned that larger demand shocks are more anticipated and thus their price impact is incorporated earlier. To test for this hypothesis, we alter regression (2) to use $r_{n,t-h\to t} = \prod_{s=0}^{h}(1+r_{n,t-s})-1$ with h as the lookback window, rather than $r_{n,t}$, as the dependent variable. We compute Newey-West standard errors with $h+1$ lags. We also lag the controls appropriately so that they are available as of time $t-h$ to avoid any lookahead bias. The results are reported in Table B.5. Columns (1)–(3), (4)–(6), and (7)–(9) present results for BMI, FIT, and OFI, respectively. We present regression results with lookback window $h = 1$, 2, and 4.

The BMI results are consistent with there being anticipation effects, as the estimated price

multipliers are larger than those in Table 2. However, if anything, the anticipation appears stronger for *smaller* shocks, and the pairwise multiplier comparisons in Panel B continue to indicate that larger demand shocks have smaller multipliers.[3]

The FIT results are also consistent with there being anticipation effects, as well as possibly fund flows chasing past returns which leads to correlations between lagged returns and flows. However, the multipliers differences in Panel B continue to show significant differences, suggesting that our results are not explained by anticipation effects. The OFI-based estimates are also consistent with there being some anticipation effects, but the multiplier differences continue to be statistically and economically significant.

Panel A: price impact regressions									
	Dependent variable: stock return $r_{n,t-h\to t}$								
	BMI			FIT			OFI		
$h=$	1	2	4	1	2	4	1	2	4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$M_{\lvert d_{n,t}\rvert<\sigma}$	2.95***	4.72***	3.12**	8.65***	11.11***	13.82***	3.86***	4.50***	5.54***
	(0.67)	(1.19)	(1.37)	(0.80)	(1.22)	(1.96)	(0.13)	(0.16)	(0.22)
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	1.06*	1.13	−0.20	6.69***	8.52***	10.95***	2.61***	3.11***	3.93***
	(0.63)	(0.91)	(0.86)	(0.50)	(0.83)	(1.47)	(0.10)	(0.12)	(0.15)
$M_{\lvert d_{n,t}\rvert>2\sigma}$	0.05	−0.15	−0.15	5.17***	6.67***	8.84***	1.15***	1.45***	2.09***
	(0.41)	(0.73)	(1.05)	(0.40)	(0.62)	(1.05)	(0.07)	(0.09)	(0.11)
Predictor controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Liquidity controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	9,088	9,056	9,011	551,912	542,021	521,884	517,042	506,380	485,361
R^2	0.174	0.212	0.237	0.084	0.089	0.093	0.124	0.128	0.131
Marginal $R^2(d_{n,t})$	0.013	0.017	0.014	0.007	0.009	0.013	0.041	0.040	0.041
Panel B: Coefficient differences									
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]} - M_{\lvert d_{n,t}\rvert<\sigma}$	−1.90**	−3.59***	−3.32**	−1.96***	−2.59***	−2.86**	−1.25***	−1.38***	−1.61***
	(0.84)	(1.35)	(1.58)	(0.56)	(0.74)	(1.12)	(0.10)	(0.12)	(0.18)
$M_{\lvert d_{n,t}\rvert>2\sigma} - M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	−1.01	−1.28	0.04	−1.52***	−1.84***	−2.12***	−1.46***	−1.66***	−1.84***
	(0.70)	(0.91)	(1.02)	(0.34)	(0.49)	(0.82)	(0.07)	(0.10)	(0.12)
$M_{\lvert d_{n,t}\rvert>2\sigma} - M_{\lvert d_{n,t}\rvert<\sigma}$	−2.90***	−4.87***	−3.28*	−3.48***	−4.43***	−4.98***	−2.71***	−3.05***	−3.45***
	(0.73)	(1.42)	(1.74)	(0.68)	(0.97)	(1.51)	(0.10)	(0.14)	(0.22)

Table B.5. Price multiplier by contemporaneous demand sizes: anticipation

This table is similar to Table 2 except that the independent variable is $r_{n,t-h\to t}$, the cumulative stock return in period $t-h$ to t, rather than the return in period t. Columns (1)–(3) report results using the ΔBMI demand shock and monthly returns. Columns (4)–(6) and (7)–(9) report results using quarterly returns and the FIT and OFI demand shocks, respectively. All specification control for 13 return-predicting characteristics and 6 liquidity proxies; those controls are described in Section 3.1. Panel B reports pairwise differences in price multipliers. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Differential reversal speeds. Following the discussion in Section 4.2, Table B.6 tests whether large shocks are associated with quicker return reversals. To this end, we estimate

[3]Extending the return window back in time introduces a correlation of the outcome variable with the May-rank-date market cap control variable, which may explain the small and negative multipliers for the larger shocks in some specifications.

Fama-MacBeth regressions of:

$$r_{n,t+1\to t+h} = \beta_0 d_{n,t} + \beta_1 d_{n,t} \times I_{|d_{n,t}|\in[\sigma_t,2\sigma_t]} + \beta_2 d_{n,t} \times I_{|d_{n,t}|>2\sigma_t} + \boldsymbol{c}_t' \boldsymbol{x}_{n,t-1} + \tau_t + \epsilon_{n,t}$$

where the dependent variable is the cumulative h-period return after realization of the demand shock. We would be concerned if, for instance, β_2 is very negative, which would indicate that the price impact of larger shocks reverts more quickly. The results for horizons h = 1,2, and 4 are reported in Table B.6. We do not find significant evidence of stronger reversals for larger shocks.

Dependent variable: future stock return $r_{n,t+1\to t+h}$									
	BMI			FIT			OFI		
$h =$	1	2	4	1	2	4	1	2	4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$d_{n,t}$	0.19	0.13	−0.19	−0.51	−1.47*	−1.92	0.04	0.10*	0.13*
	(1.48)	(2.10)	(1.94)	(0.50)	(0.88)	(1.29)	(0.05)	(0.06)	(0.08)
$d_{n,t} \cdot I_{\{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]\}}$	−0.75	−0.28	−0.84	0.23	0.69	0.67	0.03	0.12*	0.28***
	(0.88)	(1.51)	(1.91)	(0.30)	(0.54)	(0.86)	(0.04)	(0.06)	(0.10)
$d_{n,t} \cdot I_{\{\lvert d_{n,t}\rvert>2\sigma\}}$	0.28	−0.43	−0.46	0.32	1.11	1.67	0.17***	0.33***	0.65***
	(1.39)	(1.98)	(1.79)	(0.40)	(0.72)	(1.09)	(0.05)	(0.07)	(0.10)
Predictor controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Liquidity controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	8,873	8,729	8,515	544,789	528,361	496,572	516,423	505,032	482,684
R^2	0.206	0.206	0.198	0.083	0.096	0.110	0.091	0.108	0.127
Marginal $R^2(d_{n,t})$	0.015	0.014	0.014	0.002	0.002	0.003	0.003	0.004	0.005

Table B.6. Testing whether large demand shocks lead to faster price reversals

We use Fama-Macbeth regressions to investigate whether larger demand shock sizes are associated with stronger return reversals in the subsequent $h = 1, 2,$, and 4 quarters. The first three columns, next three columns, and last three columns examine the demand shocks of BMI, FIT, and OFI, respectively. The controls are the same as the full set of controls in Table 2. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

B.3 Dynamic Results

Varying lookback window. Table B.7 presents an extension of Table 4 where we vary the lookback window in constructing past cumulative demand shocks ($L = 1$ to 4). The results are robust to alternative lookback windows.

Panel A: price impact regressions								
	Dependent variable: stock return $r_{n,t}$							
$d_{n,t} =$	FIT				OFI			
	$L=1$	2	3	4	$L=1$	2	3	4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert<\sigma}$	3.95***	3.84***	3.81***	3.64***	1.79***	1.80***	1.80***	1.84***
	(0.36)	(0.35)	(0.32)	(0.32)	(0.13)	(0.14)	(0.15)	(0.15)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert\in[\sigma,2\sigma]}$	2.82***	2.77***	2.70***	2.74***	1.49***	1.50***	1.53***	1.45***
	(0.29)	(0.32)	(0.31)	(0.32)	(0.11)	(0.11)	(0.11)	(0.12)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert>2\sigma}$	1.72***	1.77***	1.76***	1.89***	1.12***	1.19***	1.21***	1.20***
	(0.30)	(0.29)	(0.30)	(0.29)	(0.09)	(0.11)	(0.13)	(0.13)
Predictor controls	Y	Y	Y	Y	Y	Y	Y	Y
Liquidity controls	Y	Y	Y	Y	Y	Y	Y	Y
Interacted: predictors	Y	Y	Y	Y	Y	Y	Y	Y
Interacted: liquidity	Y	Y	Y	Y	Y	Y	Y	Y
Obs	538,398	517,033	497,185	478,324	496,157	469,567	445,657	423,696
R^2	0.089	0.090	0.092	0.094	0.147	0.150	0.153	0.154
Panel B: Coefficient differences								
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert\in[\sigma,2\sigma]} - M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert<\sigma}$	−1.13***	−1.08***	−1.11***	−0.90***	−0.30***	−0.30***	−0.27***	−0.39***
	(0.25)	(0.26)	(0.26)	(0.25)	(0.06)	(0.06)	(0.06)	(0.05)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert>2\sigma} - M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert\in[\sigma,2\sigma]}$	−1.10***	−0.99***	−0.95***	−0.85***	−0.36***	−0.31***	−0.33***	−0.25***
	(0.22)	(0.24)	(0.24)	(0.23)	(0.07)	(0.05)	(0.06)	(0.06)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert>2\sigma} - M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert<\sigma}$	−2.22***	−2.07***	−2.06***	−1.75***	−0.66***	−0.61***	−0.60***	−0.64***
	(0.33)	(0.34)	(0.30)	(0.31)	(0.07)	(0.08)	(0.08)	(0.08)

Table B.7. Price multipliers by past cumulative demand sizes: varying lookback

This table is the same as columns (4) and (8) of Table 4 except that we vary the lookback window L (quarters). Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Combining dynamic and static effects. To investigate the marginal contribution of contemporaneous and lagged demand shock sizes in determining multipliers, we estimate cross-sectional regressions combining the independent variables from Table 2 and Table 4:

$$r_{n,t} = \sum_{\text{bin } b} M_{b,t} \cdot I_{|d_{n,t}|\in b} \cdot d_{n,t} + \sum_{\text{bin } B} M_{B,t} \cdot I_{|\sum_{l=1}^{L} d_{n,t-l}|\in B} \cdot d_{n,t} + \mathbf{c}_t' \mathbf{x}_{n,t-1} + \tau_t + \epsilon_{n,t} \quad \text{(B.1)}$$

where b and B represent bins based on contemporaneous and lagged (with $L = 4$, as in Table 4) demand shock sizes, respectively. In presenting the results in Table B.8, we omit the regressor interacted with the largest contemporaneous shock size bin to avoid perfect

collinearity.[4] The columns differ in their use of controls: columns (1)–(2) and (5)–(6) include the controls from Table 2, while columns (3)–(4) and (7)–(8) further include the controls from Table 3.

Overall, the results are broadly consistent with both contemporaneous and lagged demand shock sizes influencing price multipliers. For FIT, we observe that lagged demand shock sizes have a strong influence on endogenous multipliers, as shown by the last three rows in Panel B. After controlling for lagged shock sizes, contemporaneous demand shock sizes remain relevant but exhibit weaker statistical significance, as seen in the first two rows of Panel A and the first row of Panel B. Conversely, for OFI, contemporaneous demand shock sizes appear to drive multipliers more strongly, with lagged demand shock sizes also contributing but with weaker statistical significance.

[4]Note that if we do not omit any bins, $\sum_b I_{|d_{n,t}|\in b} \cdot d_{n,t} = \sum_B I_{|\sum_{l=1}^{L} d_{n,t-l}|\in B} \cdot d_{n,t} = d_{n,t}$, which induces perfect collinearity in the regression.

Panel A: price impact regressions								
	Dependent variable: stock return $r_{n,t}$							
$d_{n,t}$ =	FIT				OFI			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$M_{\lvert d_{n,t}\rvert<\sigma}$	0.67*	0.65*	0.63*	0.53	1.87***	2.09***	2.21***	2.51***
	(0.41)	(0.35)	(0.35)	(0.35)	(0.12)	(0.12)	(0.11)	(0.13)
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]}$	0.03	0.14	0.20	0.09	1.03***	1.13***	1.19***	1.36***
	(0.25)	(0.23)	(0.24)	(0.23)	(0.10)	(0.08)	(0.09)	(0.11)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert<\sigma}$	3.78***	3.65***	3.51***	3.54***	0.93***	0.96***	0.85***	0.66***
	(0.34)	(0.34)	(0.33)	(0.30)	(0.15)	(0.15)	(0.15)	(0.19)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert\in[\sigma,2\sigma]}$	2.76***	2.78***	2.64***	2.70***	0.82***	0.94***	0.71***	0.52***
	(0.34)	(0.34)	(0.32)	(0.31)	(0.11)	(0.12)	(0.11)	(0.15)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert>2\sigma}$	2.00***	1.98***	1.74***	1.83***	0.80***	0.93***	0.69***	0.46***
	(0.31)	(0.31)	(0.28)	(0.29)	(0.12)	(0.11)	(0.09)	(0.17)
Predictor controls	Y	Y	Y	Y	Y	Y	Y	Y
Liquidity controls	N	Y	Y	Y	N	Y	Y	Y
Interacted: predictors	N	N	Y	Y	N	N	Y	Y
Interacted: liquidity	N	N	N	Y	N	N	N	Y
Obs	478,324	478,324	478,324	478,324	423,696	423,696	423,696	423,696
R^2	0.071	0.086	0.092	0.095	0.118	0.138	0.154	0.167
Panel B: Coefficient differences								
$M_{\lvert d_{n,t}\rvert\in[\sigma,2\sigma]} - M_{\lvert d_{n,t}\rvert<\sigma}$	−0.64*	−0.50	−0.43	−0.44	−0.83***	−0.96***	−1.03***	−1.14***
	(0.33)	(0.33)	(0.32)	(0.32)	(0.10)	(0.11)	(0.11)	(0.10)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert\in[\sigma,2\sigma]} - M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert<\sigma}$	−1.01***	−0.87***	−0.86***	−0.84***	−0.11	−0.03	−0.14*	−0.13**
	(0.27)	(0.28)	(0.26)	(0.25)	(0.11)	(0.09)	(0.08)	(0.06)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert>2\sigma} - M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert\in[\sigma,2\sigma]}$	−0.77***	−0.80***	−0.90***	−0.87***	−0.03	−0.01	−0.02	−0.07
	(0.25)	(0.24)	(0.21)	(0.22)	(0.09)	(0.09)	(0.07)	(0.07)
$M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert>2\sigma} - M_{\lvert\sum_{l=1}^{L} d_{n,t-l}\rvert<\sigma}$	−1.78***	−1.67***	−1.76***	−1.72***	−0.14	−0.04	−0.16	−0.20**
	(0.35)	(0.33)	(0.31)	(0.31)	(0.18)	(0.15)	(0.11)	(0.08)

Table B.8. Price multipliers by past and contemporaneous demand sizes

Panel A reports results for Fama-MacBeth regression (B.1). In each period, we split the sample into three bins b (B) by comparing the absolute value of $d_{n,t}$ ($\sum_{l=1}^{L} d_{n,t-l}$) against their respective cross-sectional standard deviations, where $L = 4$. Columns (1)-(4) and (5)-(8) report results based on quarterly FIT and OFI demand shocks, respectively, and the columns differ by the controls included. Panel B reports pairwise differences in estimated coefficients. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

C Supplements to Section 5

C.1 Price Dynamics Under Demand Curve (12)

We use a first order Taylor expansion to show that the demand curve (12)[5]

$$\begin{aligned}\Delta q_{i,n,t} &= -\left(\zeta_1 + \zeta_2 \tilde{P}_{n,t}\right) \Delta p_{n,t} + u_{i,n,t} \\ \tilde{P}_{n,t} &= \left|\sum_{l=1}^{L} \Delta p_{n,t-l}\right| - \mathbb{E}_t^{CX}\left[\left|\sum_{l=1}^{L} \Delta p_{n,t-l}\right|\right],\end{aligned}$$

can give rise to a parametric version of the price dynamics we find in the dynamic price impact regression (9) from Section 4.3:

$$\Delta p_{n,t} = b_{1,t} \cdot d_{n,t} + b_{2,t} \cdot d_{n,t} \left|\sum_{l=1}^{L} d_{n,t-l}\right| + \epsilon_{n,t}.$$

To see this, first aggregate the demand curve across investors i using ownership-share weights $S_{i,n,t-1}$:

$$\Delta q_{S,n,t} = -\left(\zeta_1 + \zeta_2 \tilde{P}_{n,t}\right) \Delta p_{n,t} + u_{S,n,t},$$

and impose market clearing

$$\begin{aligned} 0 &= \Delta q_{S,n,t} \\ \leftrightarrow \Delta p_{n,t} &= \frac{1}{\zeta_1 + \zeta_2 \tilde{P}_{n,t}} u_{S,n,t} \\ &= \frac{1}{\underbrace{\zeta_1 - \zeta_2 \mathbb{E}_t^{CX}\left[\left|\sum_{l=1}^{L} \Delta p_{n,t-l}\right|\right]}_{\equiv \zeta_{1,t}} + \zeta_2 \left|\sum_{l=1}^{L} \Delta p_{n,t-l}\right|} u_{S,n,t} \\ &= \frac{1}{\zeta_{1,t} + \zeta_2 \left|\sum_{l=1}^{L} \Delta p_{n,t-l}\right|} u_{S,n,t}. \end{aligned}$$

[5] For notational simplicity, we remove common demand shocks, although their presence would not alter the argument.

Introduce bookkeeping parameter $\eta \in [0,1]$ that scales past shocks[6] and let

$$x_{n,t}(\eta) = \left| \sum_{l=1}^{L} \Delta p_{n,t-l}(\eta) \right|$$

where

$$\begin{aligned} \Delta p_{n,t}(\eta) &= \frac{1}{\zeta_{1,t} + \zeta_2 x_{n,t}(\eta)} u_{S,n,t} \\ \Delta p_{n,t-l}(\eta) &= \frac{\eta}{\zeta_{1,t} + \zeta_2 x_{n,t-l}(\eta)} u_{S,n,t-l}, \forall l > 0. \end{aligned}$$

Now take the first-order approximation of $\Delta p_{n,t-l}(\eta)$ around $x_{n,t-l}(\eta) = 0$:

$$\Delta p_{n,t-l}(\eta) = \frac{\eta}{\zeta_{1,t}} u_{S,n,t-l} - \underbrace{\frac{\zeta_2}{\zeta_{1,t}^2} \cdot \eta \cdot u_{S,n,t-l} x_{n,t-l}(\eta)}_{=O(\eta^2)} + O\left(\eta^3\right). \tag{C.1}$$

So[7]

$$\begin{aligned} x_{n,t}(\eta) &= \left| \sum_{l=1}^{L} \Delta p_{n,t-l}(\eta) \right| \\ &= \left| \sum_{l=1}^{L} \frac{\eta}{\zeta_{1,t}} u_{S,n,t-l} \right| + O\left(\eta^2\right). \end{aligned}$$

[6] We introduce η because the absolute value function $\left|\sum_{l=1}^{L} \Delta p_{n,t-l}\right|$ is non-differentiable at $\left|\sum_{l=1}^{L} \Delta p_{n,t-l}\right| = 0$, preventing direct Taylor expansion. The parameter $\eta \in [0,1]$ scales the magnitude of past shocks, creating a smooth approximation path.

[7] Formally, since $||a| - |b|| \leq |a - b|$ we have

$$\begin{aligned} \left| x_{n,t}(\eta) - \left| \sum_{l=1}^{L} \frac{\eta}{\zeta_{1,t}} u_{S,n,t-l} \right| \right| &= \left| \left| \sum_{l=1}^{L} \Delta p_{n,t-l}(\eta) \right| - \left| \sum_{l=1}^{L} \frac{\eta}{\zeta_{1,t}} u_{S,n,t-l} \right| \right| \\ &\leq \left| \sum_{l=1}^{L} \Delta p_{n,t-l}(\eta) - \sum_{l=1}^{L} \frac{\eta}{\zeta_{1,t}} u_{S,n,t-l} \right| \\ &= O(\eta^2), \end{aligned}$$

where the last line follows from (C.1).

Now take the first-order approximation of $\Delta p_{n,t}(\eta)$ around $x_{n,t}(\eta) = 0$:

$$\begin{aligned}\Delta p_{n,t} &\approx \frac{1}{\zeta_{1,t}} u_{S,n,t} - \frac{\zeta_2}{\zeta_{1,t}^2} u_{S,n,t} x_{n,t}(\eta) + O\left([x_{n,t}(\eta)]^2\right). \\ &= \frac{1}{\zeta_{1,t}} u_{S,n,t} - \frac{\zeta_2}{\zeta_{1,t}^2} u_{S,n,t} \left| \sum_{l=1}^{L} \frac{\eta}{\zeta_{1,t}} u_{S,n,t-l} \right| + O\left(\eta^2\right),\end{aligned}$$

where the second line follows since $x_{n,t}(\eta) = O(\eta)$. Dropping all terms of order $O(\eta^2)$ or higher and setting $\eta = 1$ yields

$$\begin{aligned}\Delta p_{n,t} &= \frac{1}{\zeta_{1,t}} u_{S,n,t} - \frac{\zeta_2}{\zeta_{1,t}^3} u_{S,n,t} \left| \sum_{l=1}^{L} u_{S,n,t-l} \right| \\ &\equiv b_{1,t} \cdot d_{n,t} + b_{2,t} \cdot d_{n,t} \left| \sum_{l=1}^{L} d_{n,t-l} \right|,\end{aligned}$$

as desired if we set $d_{n,t} \equiv u_{S,n,t}$.

C.2 Estimation Details

C.2.1 Setup

We seek to estimate the following nonlinear demand curve

$$\begin{aligned}\Delta q_{i,n,t} &= -\underbrace{\left(\zeta_{1,i,t} + \zeta_{2,i,t} \tilde{P}_{n,t}\right)}_{\text{Price Elasticity of Demand}} \Delta p_{n,t} + \underbrace{\boldsymbol{\lambda}_{i,t}^{'} \boldsymbol{\eta}_{n,t} + u_{i,n,t}}_{\text{Demand Shock}} \\ \tilde{P}_{n,t} &= \left| \sum_{l=1}^{L} \Delta p_{n,t-l} \right| - \mathbb{E}_t^{CX} \left[\left| \sum_{l=1}^{L} \Delta p_{n,t-l} \right| \right],\end{aligned}$$

$\Delta q_{i,n,t}$ is the percentage change in quantity of shares demanded by investor i for stock n in quarter t. $\Delta p_{n,t}$ is percentage price change for stock n in quarter t. L is the number of lagged quarterly price changes to consider ($L = 4$ in the baseline analysis).

To increase statistical power, we parameterize $\zeta_{1,i,t}$ and $\zeta_{2,i,t}$ as linear functions of each investor's (demeaned) active share:

$$\begin{aligned}\zeta_{k,i,t} &= \zeta_{k,0,t} + \zeta_{k,\text{Active Share},t} \left(\text{Active Share}_{i,t-1-L} - \mathbb{E}_{t-1-L}^{CX}\left[\text{Active Share}_{i,t-1-L}\right]\right) \\ &= \boldsymbol{\zeta}_{k,t}^{'} \boldsymbol{X}_{i,t}\end{aligned}$$

where $\boldsymbol{X}_{i,t} = [1, \text{Active Share}_{i,t-1-L} - \mathbb{E}_{t-1-L}^{CX}[\text{Active Share}_{i,t-1-L}]]$.

The first step is to residualize $\Delta q_{i,n,t}$, $\Delta p_{n,t}$, and $\left|\sum_{l=1}^{L} \Delta p_{n,t-l}\right|$ with respect to stock characteristics $\boldsymbol{\eta}_{n,t}$. Note that by the Frisch-Waugh-Lovell theorem, this residualization is equivalent to controlling for $\boldsymbol{\eta}_{n,t}$. $\boldsymbol{\eta}_{n,t}$ contains both observed characteristics (as well as investor-quarter fixed effects) and latent characteristics estimated using principal component analysis on the within-quarter investor $\times$ stock panel of $\Delta q_{i,n,t}$. Since this panel is unbalanced and the residual demand shocks are assumed to be heteroskedastic, we estimate the latent characteristics using the deflated heteroskedastic PCA methodology of Zhou and Chen (2025). Let $\check{q}_{i,n,t}$, $\check{\boldsymbol{p}}_{n,t}$, and $\check{\boldsymbol{P}}_{n,t}$ be the residualized counterparts of $\Delta q_{i,n,t}$, $\Delta p_{n,t} \cdot \boldsymbol{X}_{i,t}$, and $\Delta p_{n,t} \cdot \tilde{P}_{n,t} \cdot \boldsymbol{X}_{i,t}$. So we have the following residualized demand curve

$$\check{q}_{i,n,t} = -\boldsymbol{\zeta}_{1,t}^{'} \check{\boldsymbol{p}}_{n,t} - \boldsymbol{\zeta}_{2,t}^{'} \check{\boldsymbol{P}}_{n,t} + u_{i,n,t}. \tag{C.2}$$

The identifying assumption of GIV in our setting is that the residual demand shocks $u_{i,n,t}$ are conditionally independent across investors both within and across quarters:

$$\forall i \neq j, \forall t, \forall l = 1, \ldots, L : \mathbb{E}_t^{CX}[u_{i,n,t} \cdot u_{j,n,t-l}] = 0.$$

From this conditional independence assumption, if we let

$$u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) = \check{q}_{i,n,t} + \boldsymbol{\zeta}_{1,t}^{'} \check{\boldsymbol{p}}_{n,t} + \boldsymbol{\zeta}_{2,t}^{'} \check{\boldsymbol{P}}_{n,t} \tag{C.3}$$

then we obtain the following two sets of moment conditions that allow us to identify $\boldsymbol{\zeta}_{1,t}$ and $\boldsymbol{\zeta}_{2,t}, \forall i, t$:

$$\begin{aligned} \mathbf{0} &= \mathbb{E}_t^{CX}\left[\boldsymbol{X}_{i,t} \cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) \cdot \underbrace{\sum_{j \neq i} S_{j,n,t-1} u_{j,n,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right)}_{\equiv z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right)}\right] \\ &= \mathbb{E}_t^{CX}\left[\boldsymbol{X}_{i,t} \cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) \cdot z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right)\right] \end{aligned} \tag{C.4}$$

and $\forall i, t$:

$$
\begin{aligned}
\mathbf{0} &= \mathbb{E}_t^{CX}\left[\boldsymbol{X}_{i,t}\cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\cdot\left(\left|\underbrace{\sum_{l=1}^{L}\sum_{j\neq i}S_{j,n,t-1-l}u_{j,n,t-l}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)}_{\equiv z_{i,n,t-l}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)}\right| - \mathbb{E}_t^{CX}\left[\left|\underbrace{\sum_{l=1}^{L}\sum_{j\neq i}S_{j,n,t-1-l}u_{j,n,t-l}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)}_{\equiv z_{i,n,t-l}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)}\right|\right]\right)\cdot\underbrace{\sum_{j\neq i}S_{j,n,t-1}u_{j,n,t}}_{\equiv z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)}\right] \\
&= \mathbb{E}_t^{CX}\left[\boldsymbol{X}_{i,t}\cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\cdot\underbrace{\left(\left|\sum_{l=1}^{L}z_{i,n,t-l}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\right| - \mathbb{E}_t^{CX}\left[\left|\sum_{l=1}^{L}z_{i,n,t-l}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\right|\right]\right)}_{\equiv\tilde{Z}_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)}\cdot z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\right] \\
&= \mathbb{E}_t^{CX}\left[\boldsymbol{X}_{i,t}\cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\cdot\tilde{Z}_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\cdot z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\right] \qquad \text{(C.5)}
\end{aligned}
$$

All expectations are taken in the cross section of stocks n. Note that the constructed instruments $z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)$ and $\tilde{Z}_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)$ depend on the elasticity parameters $\boldsymbol{\zeta}_{1,t}$ and $\boldsymbol{\zeta}_{2,t}$ because knowledge of these parameters is required to recover the residual demand shocks from the (residualized) equilibrium quantity changes (as in (C.3)).

Note that (C.4) and (C.5) give us an overidentified system. If I is the total number of investors, then there are 4 parameters to identify (since each $\boldsymbol{\zeta}_{k,t}$ has two elements) and $4I$ moment conditions. Chaudhary et al. (2024) prove that power is maximized by using a GMM weighting matrix that places more weight on the moment conditions of investors with less volatile residual demand shocks. Thus, following Chaudhary et al. (2024), we take the precision weighted averages of moment conditions (C.4) and (C.5) to obtain an exactly identified system:

$$
\forall t: \mathbf{0} = \sum_i w_{i,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\cdot\mathbb{E}_t^{CX}\left[\boldsymbol{X}_{i,t}\cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\cdot z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\right] \qquad \text{(C.6)}
$$

$$
\mathbf{0} = \sum_i w_{i,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\cdot\mathbb{E}_t^{CX}\left[\boldsymbol{X}_{i,t}\cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\cdot\tilde{Z}_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\cdot z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\right] \qquad \text{(C.7)}
$$

$$
w_{i,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right) = \frac{\left(\mathbb{V}_t^{CX}[u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)]\right)^{-1}}{\sum_j\left(\mathbb{V}_t^{CX}[u_{j,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)]\right)^{-1}} \qquad \text{(C.8)}
$$

Empirically, we find the minimizers of the following objective function for each period t

$$\min_{\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}} \sum_k \left(\sum_i \hat{w}_{i,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right) \cdot \hat{\mathbb{E}}_t^{CX}\left[X_{i,t,k} \cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right) \cdot z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\right] \right)^2 \tag{C.9}$$

$$+ \sum_k \left(\sum_i \hat{w}_{i,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right) \cdot \hat{\mathbb{E}}_t^{CX}\left[X_{i,t,k} \cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right) \cdot \tilde{Z}_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right) \cdot z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)\right] \right)^2, \tag{C.10}$$

where $\hat{\mathbb{E}}_t^{CX}$ indicates the empirical moments as opposed to the population moments and $X_{i,t,k}$ is the k-th element of $\boldsymbol{X}_{i,t}$. So $\hat{w}_{i,t}\left(\boldsymbol{\zeta}_{1,t},\boldsymbol{\zeta}_{2,t}\right)$ indicates the precision weight (C.8) using the empirical cross-sectional variances.

C.2.2 Implementation Details

Time-Variation. To improve numerical stability, we estimate the nonlinear demand curve (12) in rolling four-quarter windows as opposed to in each quarter. Specifically, the residualization process described in (C.2.1) is conducted in each quarter, but then the residual demand curve (C.2) is estimated in the four-quarter window.

Leave-One-Type Out. We compute moment conditions (C.4) and (C.5) for investor i and all other investors j of a different 13F institution type (banks, investment advisors, insurance companies, mutual funds, pension funds, and other, as well as for the residual "household" sector that includes direct holdings by households as well as by non-13F institutions). That is, we construct the instruments $z_{i,n,t}$ and $\tilde{Z}_{i,n,t}$ in a leave-one-type out fashion as opposed to a leave-one-investor out fashion.

Removing Latent Characteristics. The baseline analysis removes five latent stock characteristics using PCA from each within-quarter investor $\times$ stock panel . Appendix C.3 presents similar results when removing alternative numbers of latent characteristics.

Unbalanced Panel. Empirically, the within-quarter investor $\times$ stock panel is highly unbalanced: most investors don't hold most stocks. Thus, to increase power we replace the precision weights in (C.8) with

$$w_{i,t}\left(\boldsymbol{a}_t,\boldsymbol{b}_t\right) = \frac{\left(\mathbb{V}_t^{CX}[u_{i,n,t}]\right)^{-1}}{\sum_k \left(\mathbb{V}_t^{CX}[u_{k,n,t}]\right)^{-1}} \cdot N_{i,t}$$

where $N_{i,t}$ is the number of stocks held by investor i in quarter t. That is, we put more weight on the moment conditions of investors with more holdings.

To further address the unbalanced nature of the panel, we retain in the estimation in each quarter stocks that are held by at least 60 investors and investors that hold at least 20 stocks.

Root Selection. In some quarters, the GMM minimization problem (C.10) has multiple roots — multiple solutions $\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right)$ that set the minimized value of the objective function exactly equal to zero. This situation can arise since we are solving a nonlinear, exactly-identified system of equations, so there is no guarantee of a unique solution. In these situations, we use the following approach to select among these multiple roots:

1. Let $\mathcal{S}$ be the set of candidate roots. Let $\mathcal{S}_+$ and $\mathcal{S}_-$ be the subsets of candidate roots for which $\zeta_{1,0} > 0$ and $\zeta_{1,0} \leq 0$, respectively.

2. If only one candidate roots has $\zeta_{1,0} > 0$ (i.e., $|\mathcal{S}_+| = 1$) use that root. Selecting roots with $\zeta_{1,0} > 0$ imposes the economic prior that, on average, demand curves should slope down.

3. Otherwise, if there is more than one candidate root with $\zeta_{1,0} > 0$ (i.e., $|\mathcal{S}_+| > 1$), then we use overidentifying restrictions to select among these. While our baseline estimation solves an exactly-identified GMM problem, we can expand the problem to be overidentified using the following moment conditions:

$$\forall t : \mathbf{0} = \boldsymbol{g}_1\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) \equiv \sum_i w_{i,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) \cdot \mathbb{E}_t^{CX}\left[\boldsymbol{X}_{i,t} \cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) \cdot z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right)\right] \tag{C.11}$$

$$\begin{aligned}\forall t, l : \mathbf{0} = \boldsymbol{g}_{2,l}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) \equiv \sum_i w_{i,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) \cdot \\ \mathbb{E}_t^{CX}\left[\boldsymbol{X}_{i,t} \cdot u_{i,n,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) \cdot \left(\operatorname{sign}\left(\sum_{l'=1}^{L} z_{i,n,t-l'}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right)\right) z_{i,n,t-l}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) - \frac{1}{L} \cdot \mathbb{E}_t^{CX}\left[\left|\sum_{l=1}^{L} z_{i,n,t-l}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right)\right|\right]\right) \cdot z_{i,n,t}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right)\right]\end{aligned} \tag{C.12}$$

That is, we keep moment condition (C.6) as is and split up moment condition (C.7) into L individual moment conditions (C.12) (that all sum to (C.7)). Letting

$$\boldsymbol{g}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right) = \left[\boldsymbol{g}_1\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right), \boldsymbol{g}_{2,1}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right), \ldots, \boldsymbol{g}_{2,L}\left(\boldsymbol{\zeta}_{1,t}, \boldsymbol{\zeta}_{2,t}\right)\right]$$

we choose the candidate root that minimizes this J statistic

$$J\left(\zeta_{1,t},\zeta_{2,t}\right)=\boldsymbol{g}\left(\zeta_{1,t},\zeta_{2,t}\right)^{'}\boldsymbol{g}\left(\zeta_{1,t},\zeta_{2,t}\right). \tag{C.13}$$

That is, we chose the candidate root that best fits the expanded set of moment conditions (C.11) and (C.12) using the identity weighting matrix. This approach is analogous to the Sargan-Hansen J-test (Hansen (1982)). The J-test is a test of overidentifying restrictions: given a solution that fits one set of moment conditions, how well does it fit a different ("held-out") set of moment conditions (where fit is evaluated based on the J-statistic (C.13), possibly with a more general weighting matrix). While the J-test is usually used to test if a model is correctly specified, here we use the J-statistic as a "model" selection criterion. Doing so ensures the selected solution is the most consistent with all available moment conditions (C.11) and (C.12), not only the particular linear combinations (C.6) and (C.7) used in the exactly-identified GMM problem (C.10).

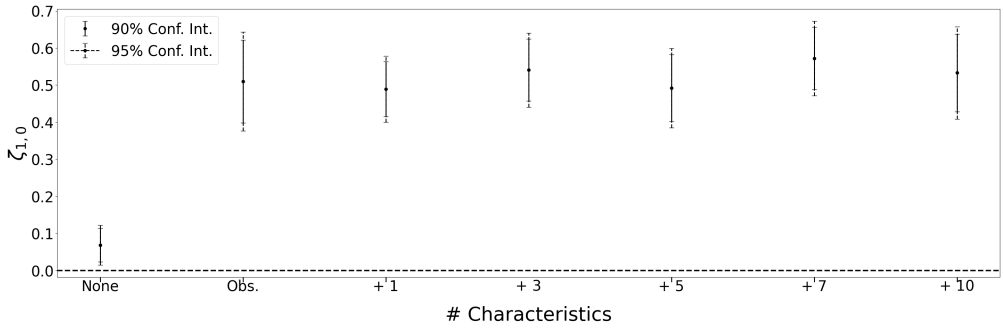
C.3 Supplemental Empirical Results

C.3.1 Alternate Numbers of Latent Factors

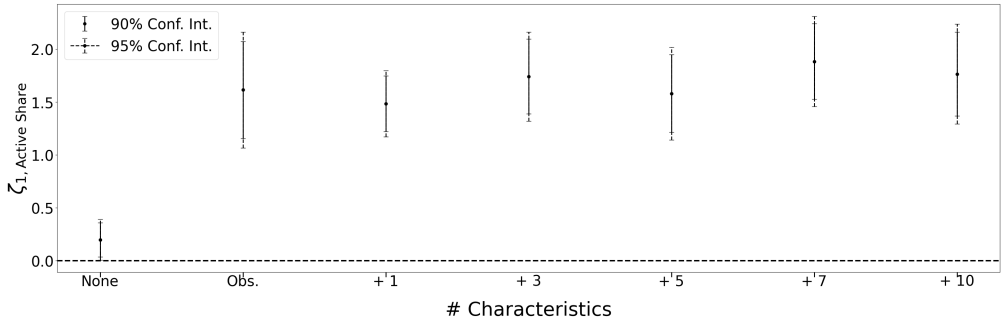
In Section 5.1.3, we extract five principal components from institutional demand. Here, we show that our results are robust to alternative number of factors. In Figure C.4, we vary the controls used and plot the resulting demand elasticity estimates. The "none" specification displays results with no observed or latent controls.

The "Obs" specification controls only for the observed return-predictor characteristics described in Section 3.1 and dummies for the Fama-French twelve industries. As expected, controlling for these characteristics raises the estimated average elasticity parameters $\zeta_{1,0}$ and $\zeta_{1,\text{Active Share}}$. Failing to control for any characteristics induces a correlation between the residual demand shocks of different investors. This correlation violates moment condition (13) and generally leads to downward bias in estimated elasticities (Chaudhary et al., 2024). Controlling for these characteristics partials out the correlation between the residual demand shocks of different investors and removes this downward bias.

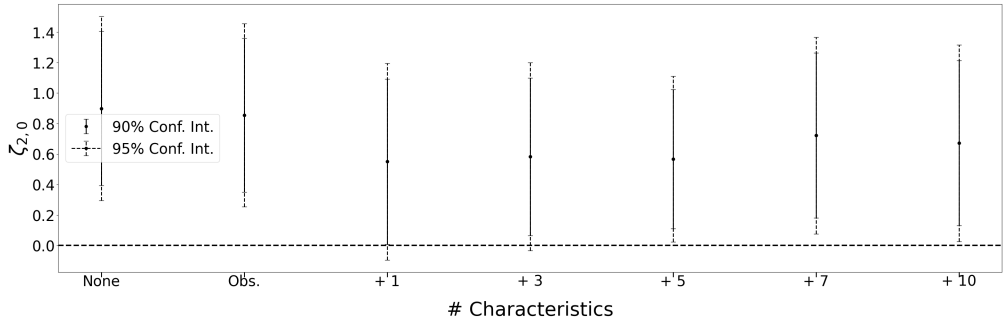
The subsequent specifications further control for latent stock-quarter characteristics extracted from PCA and vary the number of latent factors. As long as the observed characteristics are included, our results are not sensitive to the number of latent variables characteristics.



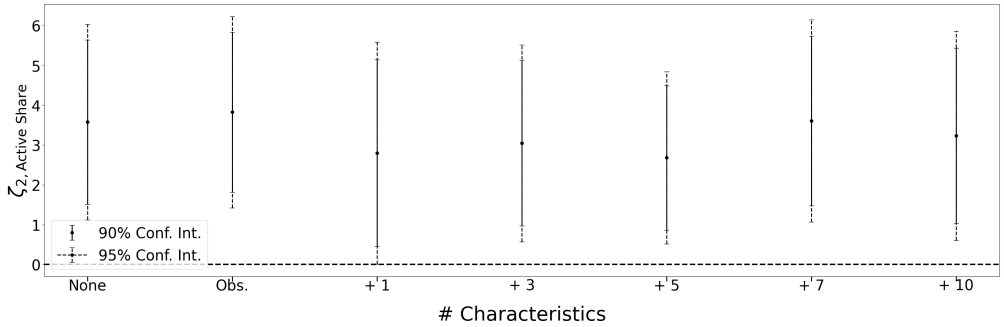
(a) $\zeta_{1,0}$



(b) $\zeta_{1,\text{Active Share}}$



(c) $\zeta_{2,0}$



(d) $\zeta_{2,\text{Active Share}}$

Figure C.4. Robustness to the number of latent factors

Panels (a) through (d) display the estimation results for alternate specifications of the non-linear demand curve (12). The "None" specification displays the estimation results when controlling for no observed or latent characteristics. The "Obs." specification controls only for the return-predicting stock characteristics described in Section 3.1 and dummies for the Fama-French twelve industries. All subsequent specifications control for both the observed stock characteristics and industry indicators, as well the number of latent stock-quarter characteristics (estimated with PCA) indicated on the x-axis.

D Proofs for Section 6

D.1 Proofs for Fixed Adjustment Cost Model in Section 6.1

Type-I investors face a fixed adjustment cost λ in $t = 1$ if they adjust from their $t = 0$ quantity Θ_0:

$$C(Q_I) = \begin{cases} 0 & , Q_I = \Theta_0 \\ \lambda & , Q_I \neq \Theta_0 \end{cases}.$$

Note that conditional on adjusting all investors (both types I and E) choose

$$Q = \frac{\mathbb{E}\left[\tilde{D} - P\right]}{\gamma \mathbb{V}\left[\tilde{D}\right]}. \tag{D.1}$$

Thus, type-I investors decides to adjust if the utility benefit of adjustment exceeds the fixed cost:

$$\begin{aligned} \underbrace{\lambda}_{\text{Cost of Adjustment}} &< \underbrace{\frac{\mathbb{E}\left[\tilde{D} - P\right]^2}{2\gamma \mathbb{V}\left[\tilde{D}\right]} - \left[\Theta_0 \mathbb{E}\left[\tilde{D} - P\right] - \frac{\gamma}{2}\Theta_0^2 \mathbb{V}\left[\tilde{D}\right]\right]}_{\text{Benefit of Adjustment}} \\ \leftrightarrow 0 &< \frac{1}{2\gamma\sigma_D^2}\mu^2 - \Theta_0\mu + \frac{\gamma}{2}\Theta_0^2\sigma_D^2 - \lambda, \text{where } \mu = \mathbb{E}\left[\tilde{D} - P\right] \\ \leftrightarrow 0 &< (\mu - \mu_+)(\mu - \mu_-). \end{aligned}$$

The adjustment thresholds are given by:

$$\begin{aligned} \mu_+ &= \Theta_0\gamma\sigma_D^2 + \sqrt{2\lambda\gamma\sigma_D^2} \\ \mu_- &= \Theta_0\gamma\sigma_D^2 - \sqrt{2\lambda\gamma\sigma_D^2}. \end{aligned}$$

So aggregate demand $(1-\epsilon)Q_I + \epsilon Q_E$ is

$$Q_{\text{agg}} = \begin{cases} \frac{\epsilon\mu}{\gamma\mathbb{V}\left[\tilde{D}\right]} + (1-\epsilon)\Theta_0 & , \mu \in [\mu_-, \mu_+] \\ \frac{\mu}{\gamma\mathbb{V}\left[\tilde{D}\right]} & , \mu \notin [\mu_-, \mu_+] \end{cases},$$

By market clearing $Q_{\text{agg}} = \Theta_1$ we have the following equilibrium price:

$$P = \begin{cases} \bar{D} - \frac{\gamma\sigma_D^2(\Theta_1 - (1-\epsilon)\Theta_0)}{\epsilon} & , \mu \in [\mu_-, \mu_+] \\ \bar{D} - \gamma\sigma_D^2\Theta_1 & , \mu \notin [\mu_-, \mu_+] \end{cases}$$

Note that

$$\begin{aligned} \mu > \mu_+ &= \Theta_0\gamma\sigma_D^2 + \sqrt{2\lambda\gamma\sigma_D^2} \\ \leftrightarrow \Theta_1 &> \Theta_0 + \frac{\sqrt{2\lambda}}{\sqrt{\gamma\sigma_D^2}}, \end{aligned}$$

and

$$\begin{aligned} \mu < \mu_- &= \Theta_0\gamma\sigma_D^2 - \sqrt{2\lambda\gamma\sigma_D^2} \\ \leftrightarrow \Theta_1 &< \Theta_0 - \frac{\sqrt{2\lambda}}{\sqrt{\gamma\sigma_D^2}}. \end{aligned}$$

Thus, we can express the adjustment thresholds in quantity space:

$$\mu \in [\mu_-, \mu_+] \leftrightarrow \Theta_1 \in \left[\Theta_0 - \frac{\sqrt{2\lambda}}{\sqrt{\gamma\sigma_D^2}}, \Theta_0 + \frac{\sqrt{2\lambda}}{\sqrt{\gamma\sigma_D^2}}\right],$$

Hence, we can write the equilibrium price as

$$P = \begin{cases} \bar{D} - \frac{\gamma\sigma_D^2(\Theta_1 - (1-\epsilon)\Theta_0)}{\epsilon} & , |\Theta_1 - \Theta_0| \leq \frac{\sqrt{2\lambda}}{\sqrt{\gamma\sigma_D^2}} \\ \bar{D} - \gamma\sigma_D^2\Theta_1 & , |\Theta_1 - \Theta_0| > \frac{\sqrt{2\lambda}}{\sqrt{\gamma\sigma_D^2}} \end{cases},$$

So the price multiplier is

$$M\left(\Theta_1 - \Theta_0\right) \equiv -\frac{P(\Theta_1) - P(\Theta_0)}{\Theta_1 - \Theta_0} = \begin{cases} \frac{\gamma\sigma_D^2}{\epsilon} & , |\Theta_1 - \Theta_0| \leq \frac{\sqrt{2\lambda}}{\sqrt{\gamma\sigma_D^2}} \\ \gamma\sigma_D^2 & , |\Theta_1 - \Theta_0| > \frac{\sqrt{2\lambda}}{\sqrt{\gamma\sigma_D^2}} \end{cases},$$

which is decreasing in the supply shock magnitude $|\Theta_1 - \Theta_0|$ since $\epsilon < 1$.

D.2 Proof of Equation (29) for Section 6.2

Proof. At $t = 0$ the investor chooses the optimal gain for $t = 1$ after observing Θ_0. She also knows the equilibrium distributions of quantity $Q_1(\mathcal{I}_1)$ (for $\mathcal{I}_1 = (s, G)$) and price P, both of which she takes as given.[8]

Rewrite the information choice problem (25) in terms of the optimal $t = 1$ quantity, which is a function of the chosen gain G:

$$\max_G \mathbb{E}_0\left[Q_1(\mathcal{I}_1)\cdot\mathbb{E}_1\left[\bar{D}-P\right]-\frac{\gamma}{2}Q_1(\mathcal{I}_1)^2\cdot\mathbb{V}_1\left[\bar{D}-P\right]\right]-C(G).$$

Note that by the envelope theorem, $\partial Q_1(\mathcal{I}_1)/\partial G = 0$. Thus, taking the first-order condition with respect to G yields

$$\begin{aligned}
0 &= \mathbb{E}_0\left[-\frac{\gamma}{2}Q_1(\mathcal{I}_1)^2\cdot\frac{\partial}{\partial G}\mathbb{V}_1\left[\bar{D}-P\right]-C'(G)\right]\\
&= \mathbb{E}_0\left[-\frac{\gamma}{2}Q_1(\mathcal{I}_1)^2\cdot\frac{\partial}{\partial G}\left[\sigma_\varepsilon^2+\sigma_\eta^2(1-G)\right]-C'(G)\right]\\
&= \mathbb{E}_0\left[\frac{\gamma}{2}Q_1(\mathcal{I}_1)^2\cdot\sigma_\eta^2-C'(G)\right]\\
&= \mathbb{E}_0\left[\frac{\gamma}{2}\Theta_0^2\cdot\sigma_\eta^2-C'(G)\right]\\
\leftrightarrow \frac{\gamma}{2}\sigma_\eta^2\left(\Theta_0^2+\sigma_\Theta^2\right) &= C'(G),
\end{aligned}$$

where the fourth equation follows by market clearing (26), and the fifth equation follows since information choice is done after observing Θ_0.

□

[8] We assume there is a continuum of identical, infinitesimal investors. Therefore, the representative investor does not internalize the impact of her information choice on price, nor does she internalize the impact of her portfolio choice on price.