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

Aditya Chaudhry and Jiacui Li

March 17, 2025

ABSTRACT

We document a new stylized fact about inelastic demand in financial markets: larger uninformed demand shocks have smaller per-unit price impact (i.e. smaller price multipliers). That is, we find total price impact is concave in the size of demand shocks. This finding reveals an important dimension of endogenous variation in price multipliers. Since many existing theories imply convex or linear price impact, our finding helps discipline potential theories of inelastic demand. Based on these insights, we propose a nonlinear asset demand system with endogenous price elasticities of demand. The estimated demand system demonstrates this concavity is quantitatively important: extrapolating local price multiplier estimates may overstate the impact of large quantity shifts on prices.

JEL classification: G11, G12

Keywords: Demand-Based Asset Pricing, Inelastic Demand, 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 Ehsan Azarmsa, Carter Davis, Espen Eckbo, Zhiyu Fu, Michael Gallmeyer, Niels Gormsen, Valentin Haddad, Lars Hansen, Paymon Khorrami, Ralph Koijen, Erik Loualiche, Stefan Nagel, Simon Oh, Anna Pavlova, Kaushik Vasudevan, and participants of the UVA McIntire Finance Brownbag.

1 Introduction

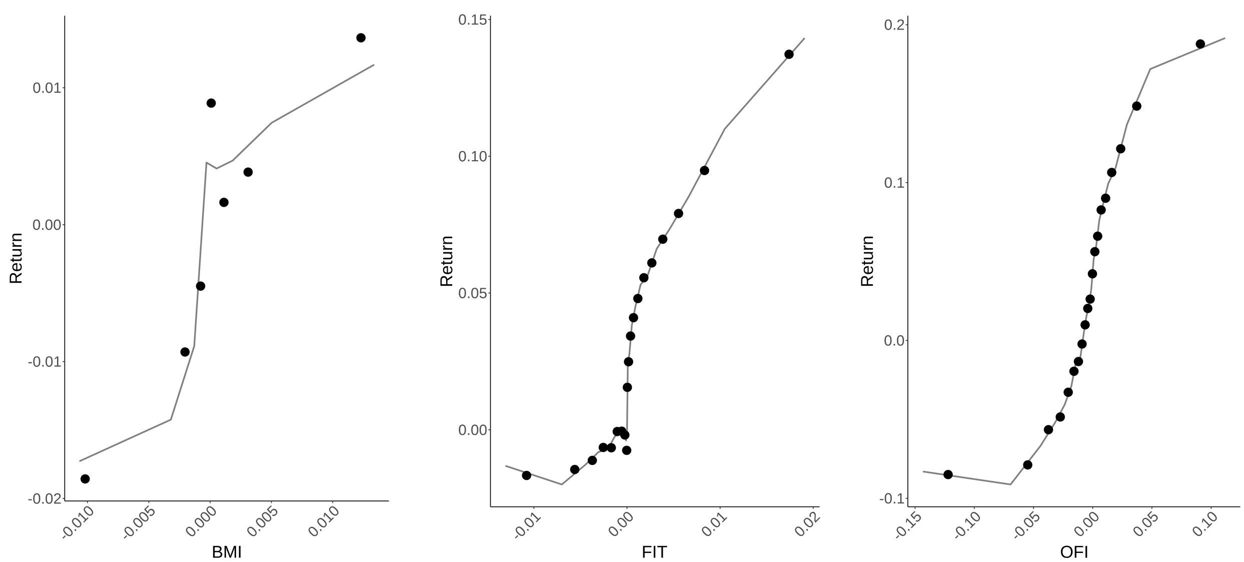
Many questions in asset pricing and macro-finance concern the impact of large shifts in quantities on asset prices. How do government asset purchases impact prices? How has the shift to passive investing impacted price efficiency? How has the rise of green investing impacted stock prices and firms' cost of capital? These questions motivate a growing literature that studies inelastic demand in asset markets and measures "price multipliers": the per-unit impact of investor demand shocks on prices (Koijen and Yogo (2019); Gabaix and Koijen (2022); Haddad et al. (2024)). Given the difficulty in finding natural experiments involving large demand shocks, to cleanly identify these multipliers, the literature has focused on plausibly exogenous but relatively small shocks. These studies find large price multipliers.

Yet a fundamental question remains unresolved: How do these price multipliers measured from small shocks relate to those that would arise for the large shocks in the motivating questions? This question persists because the true microfoundations underlying large price multipliers remains unknown. While there are many potential explanations for large multipliers, different models offer conflicting predictions about how multipliers endogenously vary with shock size. Some models suggest multipliers increase with shock size, but others predict they remain constant or decrease. Thus, empirical evidence on how price multipliers vary with shock is essential to understand how large demand shocks impact prices.

In this paper, we document a new stylized fact: stock-level price multipliers are smaller for larger demand shocks. Using three common demand shocks from the literature, we find that larger shocks have smaller per unit price impact, as displayed in Figure 1. This pattern implies that total price impact is concave in shock size.

To quantify the importance of this concavity, we propose a nonlinear asset demand system with endogenous price elasticities of demand. We find this concavity is quantitatively important for counterfactual analyses in financial markets: multipliers for small shocks (10 basis points) are three times larger than multipliers for large shocks (10 percentage points), with values of 10 and 3 respectively.

Our findings have three important implications. First, the concavity we document suggests that price multipliers for the large demand shocks in the motivating questions are likely smaller than multipliers measured from small shocks. Second, our new stylized fact sheds light on the crucial question of why markets feature inelastic demand and large price multipliers. Since many models imply convex or linear price impact, our findings suggest such models miss first-order determinants of price multipliers. Thus, our new stylized fact disciplines the set of theories that can explain large price multipliers. Third, this result provides a new empirical moment to discipline asset pricing and macro-finance models more broadly. Many of these models rely on frictions that do not generate concave price impact, which suggests other frictions may be more important in reality.



(a) (b) (c)

Figure 1. Price impact is concave
This Figure plots binscatter plots and piecewise linear regressions of average stock returns against demand shocks by bins, using the method of Cattaneo et al. (2024). The demand shocks include the benchmarking intensity (BMI) of Pavlova and Sikorskaya (2023), flow-induced trading (FIT) of Lou (2012), and order flow imbalance (OFI) in Li and Lin (2023). We include the controls from Pavlova and Sikorskaya (2023) in Plot (a).

We begin by presenting a stylized model to illustrate how different potential microfoundations for large price multipliers yield conflicting predictions for how multipliers vary with demand shock size (in Section 2). We model a representative investor who maximizes mean-variance utility subject to a general cost function that reduces his willingness to absorb exogenous supply shocks. This cost function nests many frictions including physical trading costs, psychological costs of adjustment, shadow costs of constraints, and even subjective uncertainty. Many common mechanisms imply price impact is either convex (e.g. certain convex adjustment costs (Gârleanu and Pedersen (2013); Bacchetta et al. (2023)) or leverage constraints (He and Krishnamurthy (2013); Brunnermeier and Sannikov (2014))) or linear (e.g. benchmarking or investment mandates (Petajisto (2009); Pavlova and Sikorskaya (2023); Gabaix and Koijen (2022))). We discuss what conditions on the cost function give rise to concave price impact as well as what classes of mechanisms satisfy these conditions. In particular, concavity is consistent with two broad classes of mechanisms: those in which it is less costly to take larger positions than smaller ones (e.g. fixed costs, as in Lo et al. (2004)), and those in which investors can pay to expand their risk-bearing capacity when presented with profitable opportunities (e.g. costly information acquisition as in Van Nieuwerburgh and Veldkamp (2009), Van Nieuwerburgh and Veldkamp (2010), Kacperczyk et al. (2016), or Han (2018)) or endogenous slow-moving capital (Duffie and Strulovici (2012))). We then provide empirical evidence of concave price impact. To measure price multipliers, we use three uninformed demand shocks from previous work (as detailed in Section 3). As established in previous work, these shocks are plausibly unrelated to cash flow news and so allow us to measure how prices respond to "exogenous" changes in demand unrelated to changes in firm fundamentals.

First, we use index-reconstitution induced changes in benchmarking intensity (BMI, Pavlova and Sikorskaya, 2023; Sikorskaya, 2023). BMI measures changes in the amount of benchmarked institutional capital from funds tracking the Russell indices. Each June, stocks mechanically enter and exit the Russell 1000 and 2000 based on which side of a cutoff their May market caps fell on. Thus, the flows from benchmarked funds prompted by this reconstitution and the resulting price pressure are exogenous to June cash flow news (Chang et al. (2015); Crane et al. (2016); Glossner (2019)).

Second, we use mutual fund flow-induced trading (FIT, Lou, 2012; Li, 2022; Ben-David et al., 2022; Van der Beck, 2021, 2022; Chaudhary et al., 2023; Chaudhry, 2023). Flows induce funds to do some mechanical rebalancing: funds tend to scale their preexisting holdings proportionally in response to flows. This mechanical component of the cross-sectional trading induced by flows is plausibly unrelated to cash flow news.

Third, we use order flow imbalance (OFI, Li and Lin, 2023). OFI is calculated from all executed trades in the U.S. stock market, signed as buy or sell trades using the Lee and Ready (1991) algorithm. Thus, OFI captures all trades that are executed aggressively. Li and Lin (2023) run extensive tests and do not find evidence that OFI is related to various measures of cash flow-relevant news. However, they also acknowledge that, due to the difficulty of measuring news, one cannot be fully certain about the information content of OFI.

Each of these demand shocks has its own strengths. On one hand, BMI and FIT provide shifts in asset demand that are plausibly unrelated to cash flow news. On the other hand, OFI has far greater variation than the other two shocks, which proves useful in measuring price multipliers for large shocks (OFI has a quarterly volatility of 400 basis points versus a quarterly volatility of 60 basis points for FIT and a monthly volatility of 70 basis points for BMI). Though these shocks use different sources of variation, we find consistent results across all three, which underscores the robustness of our findings.

Across all three measures, we find evidence of concave price impact (in Section 4). Price multipliers estimated at the monthly (for BMI) and quarterly (for FIT and BMI) frequencies are smaller for shocks that are larger in magnitude. This concavity is economically significant: while buying $X\%$ of shares outstanding of a stock raises price by $1.5X\%$ to $3X\%$ (depending on the demand shock) for small X, each 1% increase in the magnitude of X reduces this impact by 14 to 59 basis points.

We conduct a series of tests to rule out alternative explanations of our results. We estimate all baseline specifications across stocks within each time period. This cross-sectional approach avoids the possibility that large demand shocks might cluster in periods when price multipliers are systematically lower. We consider alternate specifications that allow price multipliers to vary cross-sectionally with stock characteristics and still find quantitatively similar degrees of concavity. This approach assuages the potential concern that large shocks might concentrate in stocks with characteristics that generally have small multipliers. We also conduct specifications in which we strip out returns from days with high news content. Doing so addresses the potential concern that our demand shocks correlate with cash flow news (which creates positive omitted variable bias in the estimated multipliers) and this correlation is weaker for larger demand shocks. The plausible exogeneity of the BMI and FIT demand shocks to cash flow news further addresses this concern.

It is important to note that even though price multipliers decrease with shock size, they remain much larger than standard asset pricing models predict.[1] Thus, our findings should be interpreted as shedding light on the endogenous variation of price multipliers, rather than arguing that multipliers are as small (and demand is as elastic) as in standard models. It is also worth clarifying that, despite the apparent similarity, our finding is different from the microstructure "square root" price impact curves at the trade or order levels (e.g. Frazzini et al., 2018). Those studies typically focus on time periods that are intraday to up to a few days, while we focus on "asset pricing" frequencies of months to quarters, a horizon at which many microstructure effects should have dissipated. Accordingly, the explanations we consider for concave price impact is also different from that given in the econophysics literature, which is based on the shape of order books and the frequency of trading (e.g.

[1] Classical asset pricing models predict uninformed demand shocks to have stock-level price multipliers on the order of 1/6000 (e.g. Petajisto, 2009).

Bouchaud et al., 2018).

Motivated by these reduced-form results, we propose a nonlinear asset demand system to study the quantitative implications of concave price impact. Building on Koijen and Yogo (2019) and Koijen et al. (2024), we model investor-level demand for stocks as a function of stock characteristics. However, departing from previous work, we assume each investor is subject to a portfolio adjustment cost that renders his price elasticity of demand endogenous. The adjustment cost function is flexible enough to allow elasticity to increase or decrease (depending on the estimated parameters) with the size of adjustment. We estimate the demand system on institutional investor holdings data from SEC Form 13F. The estimated system reveals that most investors become more elastic when making larger adjustments, which is consistent with the concave price impact we document. We use the estimated system to study how much this concavity dampens the price impact of large demand shocks. We find that multipliers for small shocks (10 basis points) are three times larger than multipliers for large shocks (10 percentage points), with values of 10 and 3 respectively.

The paper is structured as follows. Section 2 presents our stylized model. Section 3 describes the data and demand shocks we use. Section 4 presents empirical evidence of the concave price impact. Section 5 presents the nonlinear asset demand system. Section 6 concludes.

1.1 Related Literature

This paper relates to two literatures: the demand-based asset pricing literature and the literature studying equilibrium effects of frictions in asset pricing and macro-finance.

First, our paper contributes to a growing literature on demand-based asset pricing. Previous work using asset demand systems (Koijen and Yogo (2019, 2020); Huebner (2023); Haddad et al. (2024); Koijen et al. (2024)) and plausibly exogenous demand shocks (Shleifer (1986); Harris and Gurel (1986); Chang et al. (2015); Hartzmark and Solomon (2022); Li (2022); Schmickler and Tremacoldi-Rossi (2022); Pavlova and Sikorskaya (2023)) finds price elasticities of asset demand are much smaller empirically than in standard models, and so price multipliers are much higher. Broadly, this literature has not studied endogenous variation in multipliers. This paper contributes to this literature by empirically documenting that price multipliers decline as demand shock size increases. Moreover, we establish the theoretical conditions under which this pattern arises.

A particularly relevant paper is Haddad et al. (2024), which finds price elasticites of demand (and so price multipliers) vary *across stocks* depending on which other investors hold the stock. In this paper, we document a novel and different source of variation in multipliers. We argue that multipliers depend on demand shock size: larger demand shocks have smaller price multipliers.

Second, our paper relates to the asset pricing and macro-finance literatures that study real and financial effects of various market frictions. Our new stylized fact that price impact is concave provides a new moment to discipline these models because many proposed frictions imply price impact is either convex (e.g. convex adjustment costs (Gârleanu and Pedersen (2013); Bacchetta et al. (2023)) or leverage constraints (He and Krishnamurthy (2013); Brunnermeier and Sannikov (2014))) or linear (e.g. benchmarking or investment mandates (Petajisto (2009); Pavlova and Sikorskaya (2023); Gabaix and Koijen (2022))). The concave price impact we document is consistent with models in which large adjustments are relatively cheaper on a per-unit basis than small adjustments (e.g. fixed costs (Lo et al. (2004)), and those in which investors can pay to expand their risk-bearing capacity when presented with profitable opportunities (e.g. costly information acquisition (Van Nieuwerburgh and Veldkamp (2009, 2010); Kacperczyk et al. (2016); Han (2018)) or endogenous slow-moving capital (Duffie and Strulovici (2012))).

Additionally, the industry-oriented microstructure literature documents that price impacts appear to follow "square-root laws" at higher frequencies at the trade or order levels (e.g. Tóth et al., 2011). In the econophysics literature, the microfoundation for this effect is based on microstructure arguments(e.g. Alfonsi et al., 2010; Gatheral, 2010; Donier et al., 2015). For instance, Donier et al. (2015) derives square root price impact under the assumption that the order book is locally linear. Bouchaud et al. (2018) (chapter 18) argues that price impact tends to be concave in markets that trade more frequently, while markets that trade infrequently should exhibit patterns closer to linear price impact. We believe our work is thus quite different as we focus on monthly to quarterly time scales, and no previous work has established whether concavity in microstructure settings extend to lower frequencies and for much larger demand shocks. More importantly, because we study frequencies at which many microstructure effects have arguably dissipated, the possible explanations for our findings are also different from those in the microstructure literature.

2 Theoretical Framework

In this section, we present a stylized model to illustrate the theoretical ambiguity of how price multipliers vary across demand shock size. We discuss under what conditions concave price impact can arise, and which mechanisms satisfy these conditions.

2.1 Setup

There are two periods: $t = 1, 2$.

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

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

At time $t = 1$, the asset has exogenous, stochastic supply $\Theta \sim N(0, \sigma_\Theta^2)$, which should be interpreted as the *residual* supply the representative investor faces: the total fixed, positive net supply minus the exogenous demand shocks of some noise traders. The exogenous risk-free rate is normalized to zero.

Agents and Preferences: There is a representative investor who chooses Q, the number of risky shares, at $t = 1$ to maximize mean-variance utility over $t = 2$ (i.e. terminal) wealth subject to a cost $C(Q, P)$:

$$\max_{Q} \mathbb{E}\left[Q\left(\tilde{D} - P\right)\right] - \frac{\gamma}{2}\mathbb{V}\left[Q\left(\tilde{D} - P\right)\right] - C(Q, P) \tag{1}$$

The cost function can depend both on the number of shares held (e.g. as in models with adjustment or entry costs) and on the price (e.g. as in models with leverage constraints). The properties of $C(Q, P)$ determine if price impact is convex, linear, or concave. Assume $\frac{\partial^2}{\partial Q^2} C(Q, P) > 0$ so the investor's objective function is strictly concave. The cost function nests many frictions including physical trading costs, psychological costs of adjustment, shadow costs of constraints, and even subjective uncertainty, as discussed in Section 2.2.

Taking the first-order condition of (1), the investor's optimal quantity demanded satisfies

$$P = \mathbb{E}\left[\tilde{D}\right] - Q \cdot \gamma\mathbb{V}\left[\tilde{D}\right] - \underbrace{\frac{\partial}{\partial Q} C(Q, P)}_{\equiv MC(Q,P)}. \tag{2}$$

$MC(Q, P)$ is the marginal cost of holding one more share of the asset.

Equilibrium: Plugging the market-clearing condition

$$Q = \Theta. \tag{3}$$

into the first-order condition (2) yields the equilibrium price

$$P = \mathbb{E}\left[\tilde{D}\right] - \Theta \cdot \gamma\mathbb{V}\left[\tilde{D}\right] - MC(\Theta, P). \tag{4}$$

Price depends on marginal cost. If the marginal cost of holding an additional share is high, the price must be low to incentivize the investor to hold the share in equilibrium.

2.2 Properties of Price Multiplier

Differentiating both sides of (4) with respect to P and rearranging yields the following expression for the price multiplier (the per-unit change in price due to a change in Θ):

$$M \equiv -\frac{\partial P}{\partial \Theta} = \frac{\gamma \mathbb{V}\left[\tilde{D}\right] + \frac{\partial}{\partial \Theta} MC\left(\Theta, P\right)}{1 + \frac{\partial}{\partial P} MC\left(\Theta, P\right)}. \tag{5}$$

Note that the multiplier is expressed as a positive number (since $\frac{\partial P}{\partial \Theta} < 0$ as increases in supply Θ lower price).

From (5), we see the multiplier M depends on how marginal cost varies with quantity and price. First, M is large if marginal cost increases with quantity held ($\frac{\partial}{\partial \Theta} MC\left(\Theta, P\right) > 0$). If holding more shares requires paying a higher marginal cost, the price must adjust (i.e. drop) to incentivize the investor to do so. This behavior arises in models with, for example, adjustment costs, uncertainty about expected returns, and benchmarking. Second, M is large if marginal cost increases as price falls ($\frac{\partial}{\partial P} MC\left(\Theta, P\right) < 0$). In this case, when supply rises and price falls, the investor's marginal cost rises, and so price must fall even more to compensate the investor. This behavior arises in models with leverage constraints where the shadow cost of the constraint is high when price is low (i.e. expected return is high) because the investor wants to take a large position but cannot.

Thus, the relationship between the multiplier M and shock size $|\Theta|$ depends on whether marginal cost grows faster for large or small shocks. We now discuss under which conditions each of these situations arises, as well as which models these conditions.

When does M Grow with Shock Size? M grows with shock size (represented by the red lines in Figures 2c and 2d) and price impact is convex in $|\Theta|$ (represented by the red lines in Figures 2e and 2f) in two situations. First, M rises with $|\Theta|$ if larger positions lead marginal cost to *increase faster* with quantity ($\frac{\partial^2}{\partial |\Theta| \partial \Theta} MC(\Theta, P) > 0$), as in red in Figure 2a.

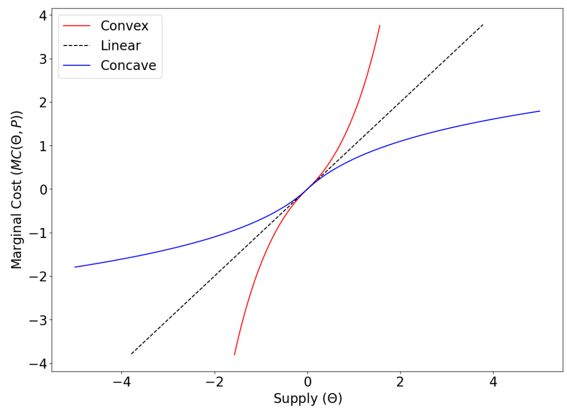
As shown in Appendix A.1, this behavior arises in certain specifications of convex adjustment costs (e.g. Gârleanu and Pedersen (2013); Bacchetta et al. (2023)): the cost of adjusting ones holdings rises with the size of the adjustment. Second, M rises with $|\Theta|$ if larger positions lead marginal cost to *increases faster* as price falls ($\frac{\partial^2}{\partial|\Theta|\partial P}MC(\Theta, P) < 0$), as in red in Figure 2b. As shown in Appendix A.2, this behavior arises with in models with leverage constraints (e.g. He and Krishnamurthy (2013); Brunnermeier and Sannikov (2014)). For small Θ the constraint does not bind and the shadow cost is zero regardless of price. But when Θ grows and the constraint binds, the shadow cost rises as price falls.

When does M Not Vary with Shock Size? M does not vary with shock size (as in black in Figures 2c and 2d) and price impact is linear (as in black in Figures 2e and 2f) if larger positions do not impact how quickly marginal cost changes with quantity or price ($\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) = \frac{\partial^2}{\partial|\Theta|\partial P}MC(\Theta, P) = 0$), as in black in Figures 2a and 2b. As shown in Appendix A.3, this behavior arises in many models with benchmarking or investment mandates (e.g. Petajisto (2009); Gabaix and Koijen (2022); Pavlova and Sikorskaya (2023)): the marginal cost of deviating from the benchmark increases at a constant rate.

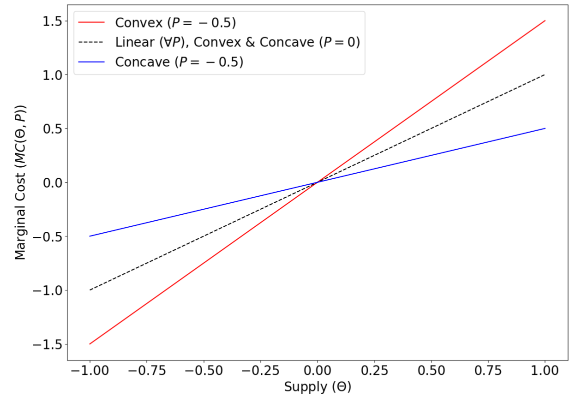
When does M Shrink with Shock Size? M shrinks with shock size (as in blue in Figures 2c and 2d) and price impact is concave (as in blue in Figures 2e and 2f) in two situations. First, M falls with $|\Theta|$ if larger positions lead marginal cost to *increase slower* with quantity ($\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) < 0$), as in blue in Figure 2a. As shown in Appendix A.4, this behavior arises in models with fixed adjustment costs: marginal cost increases significantly for small adjustments, but then falls to zero for larger adjustments past that initial hurdle (e.g. Lo et al. (2004)). Second, M falls with $|\Theta|$ if larger positions lead marginal cost to *increase slower* as price falls ($\frac{\partial^2}{\partial|\Theta|\partial P}MC(\Theta, P) > 0$), as in blue in Figure 2b. As show in Appendix A.5, this behavior arises in models where investors can endogenously expand their risk-bearing capacity, such as costly information acquisition (e.g. Van Nieuwerburgh and Veldkamp (2009, 2010); Kacperczyk et al. (2016); Han (2018)) or slow-moving capital (e.g.

Duffie and Strulovici (2012)). When price falls and raises expected return, investors can take larger positions by paying to expand their risk-bearing capacity. Doing so lowers marginal cost and reduces price impact.

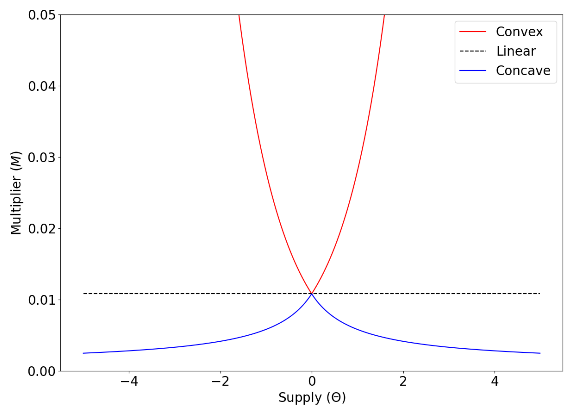
Thus, different potential microfoundations for large price multipliers yield conflicting predictions for how multipliers vary with demand shock size.



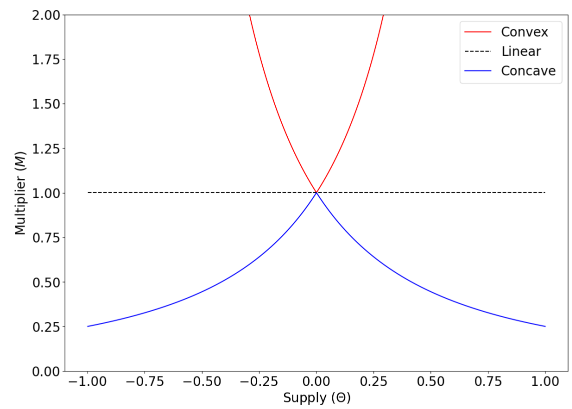
(a) Marginal cost when $\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) \lesseqqgtr 0$



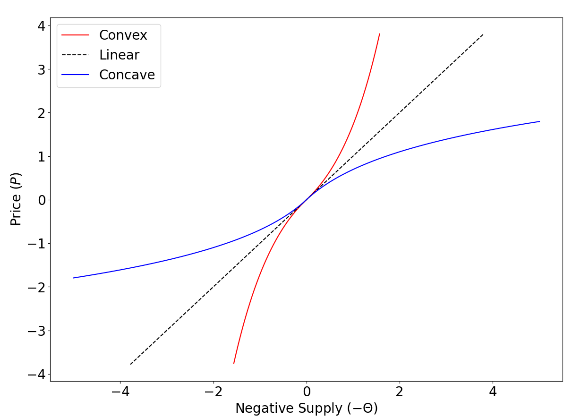
(b) Marginal cost when $\frac{\partial^2}{\partial|\Theta|\partial P}MC(\Theta, P) \lesseqqgtr 0$



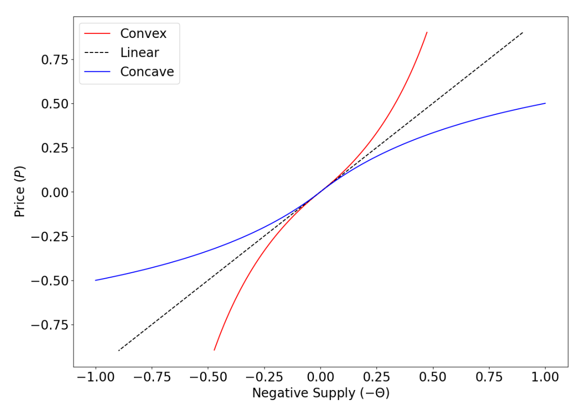
(c) Price multiplier when $\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) \lesseqqgtr 0$



(d) Price multiplier when $\frac{\partial^2}{\partial|\Theta|\partial P}MC(\Theta, P) \lesseqqgtr 0$



(e) Price when $\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) \lesseqqgtr 0$



(f) Price when $\frac{\partial^2}{\partial|\Theta|\partial P}MC(\Theta, P) \lesseqqgtr 0$

Figure 2. Panels (a), (c), and (e) plot the marginal cost $MC(Q, P)$, price multiplier M, and equilbrium price P for $\bar{D} = 0, \sigma_\epsilon = 0.02$, and $MC(Q, P) = \text{sign}(\theta)(\exp[|\theta|] - 1)$ (in red, has $\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) > 0$), $MC(Q, P) = \theta$ (in black, $\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) = 0$), and $MC(Q, P) = \text{sign}(\theta)\log(|\theta| + 1)$ (in blue, $\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) < 0$). Panels (b), (d), and (f) plot the marginal cost $MC(Q, P)$, price multiplier M, and equilbrium price P for $\bar{D} = 0, \sigma_\epsilon = 0.02$, and $MC(Q, P) = \theta(1 + |\bar{D} - P|)$ (in red, has $\frac{\partial^2}{\partial|\Theta|\partial P}MC(\Theta, P) < 0$), $MC(Q, P) = \theta$ (in black, has $\frac{\partial^2}{\partial|\Theta|\partial P}MC(\Theta, P) = 0$), and $MC(Q, P) = \theta(1 - |\bar{D} - P|)$ (in blue, has $\frac{\partial^2}{\partial|\Theta|\partial P}MC(\Theta, P) > 0$).

3 Data and Demand Shocks

To study how price multipliers vary with demand shock size, we use three demand shocks from previous work that researchers argue are largely void of cash flow information.

Bechmarking Intensity (BMI) Our first demand shock is based on index inclusion, 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 the benchmarking intensity (BMI) measure from Pavlova and Sikorskaya (2023), which provides a continuous measure of changes in demand driven by Russell index reconstitutions.

Each May, Russell ranks eligible stocks by market capitalization to determine index membership. Stocks above a specified rank enter the Russell 1000, while those below join the Russell 2000. The Russell 2000 historically attracts more benchmarked institutional capital than the Russell 1000. When stocks cross the cutoff during annual reconstitution in June, they experience institutional flows: stocks moving down to the Russell 2000 see inflows and positive returns, while those moving up to the Russell 1000 face outflows and negative returns. 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 an uninformed demand shock (Chang et al. (2015); Crane et al. (2016); Glossner (2019)).

Russell reconsitution provides demand shocks of different sizes because, as Pavlova and Sikorskaya (2023) note, these reconstitution-induced flows differ across stocks. 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 under market cap rank 200) are in the Russell Midcap Blend, Value, and Growth indices. Thus, a stock moving from the Russell 1000 Value to the Russell 2000 Value may experience a different magnitude of benchmarking demand shock than a stock moving from the Russell 1000 Growth to the

Russell 2000 Growth.[2]

The Pavlova and Sikorskaya (2023) benchmarking intensity (BMI) measure captures this variation:

$$BMI_{i,t} = \sum_{\text{Index } j} \frac{\text{Institutional AUM Benchmarked to Index } j \text{ in month } t \cdot \mathbf{1}_t \left(i \in \text{Index } j\right)}{\text{Index } j \text{ Market Value in month } t}.$$

This measure quantifies the inelastic demand for each stock from benchmarked funds. It depends on which indices a stock is part of and the proportion of each index held by benchmarked investors. BMI is calculated from thirty-four indices, including nine Russell benchmarks, covering approximately 90% of mutual fund and ETF assets.

We use June BMI changes in each year, denoted ΔBMI, for stocks in a narrow window (150 stocks in the baseline specification) around Russell reconstitution thresholds as an uninformed demand shock. Stocks with positive (negative) ΔBMI experience benchmarking inflows (outflows). While BMI is generally endogenous because index membership is, ΔBMI for stocks in this window are driven by Russell index membership changes, which are exogenous to June cash flow news conditional on the May rank-date market cap.[3]

We use the BMI and Russell index constituents 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

[2]Technically all stocks in the Blend indices are in both the Value and Growth indices, just in different proportions.

[3]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 mechanical functions of the firm size distribution. Thus, there is a "band" of market caps including stocks from the Russell 1000 and 2000. Appendix B.1 explains the Russell methodology we use to calculate these cutoffs. 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 proprietary market cap using standard databases. Doing so predicts assignment to the Russell 1000 and 2000 with high accuracy. Following previous work, we use May — not June — market caps to calculate the Russell reconstitution thresholds to avoid selection bias (e.g. Chang et al. (2015); Appel et al. (2021); Wei and Young (2021)).

spread,[4] and the banding controls from Appel et al. (2019) (an indicator for having rank-date market cap in the "band", an indicator for being in the Russell 2000 in May, and the interaction of these indicators). 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). Conditional on these variables that determine Russell 1000/2000 membership, $\Delta BMI_{i,t}$ in June is unrelated to June cash flow news.

Flow-Induced Trading (FIT) Our second demand shock is the flow-induced trading (FIT) measure from Lou (2012), which captures the stock-level trading by mutual funds in response to fund flows.

Fund flows induce uninformed stock-level trading by mutual funds, which tend to scale pre-existing holdings proportionally to ex-ante portfolio weights (Frazzini and Lamont (2008)). For example, if Apple's existing weight is 5% in a fund's portfolio, a \$1 inflow (outflow) induces the fund to increase (decrease) capital allocation of about five cents to Apple, a behavior that is documented in Coval and Stafford (2007) and Lou (2012), among others. This behavior is not only true for index funds, but also approximately true for active mutual funds and exchange-traded funds (Figure A4 in Li (2022)). This predicted mechanical component of the cross-sectional trading due to flows is uninformed.

We use the FIT instrument of Lou (2012). We first calculate the quarterly (percentage) flow to mutual fund n as

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

where $TNA_{n,t}$ and $\text{Ret}_{n,t}$ are fund n's total net assets in quarter t and return from quarter $t-1$ to t, respectively. The predicted mechanical trading by fund n in stock i due to this flow is $\text{SharesHeld}_{n,i,t-1} \cdot f_{n,t}$. Aggregating across all funds and scaling by shares outstanding

[4] Pavlova and Sikorskaya (2023) note changes in a stock's liquidity can impact both its returns (by altering the liquidity premium) and BMI. Thus, they control for Russell's proprietary float factor and the rolling average bid-ask percentage spread (to address staleness in the float factor). Lacking access to Russell's proprietary float factor, we control for the bid-ask spread.

yields:[5]

$$\text{FIT}_{i,t} = \sum_{\text{fund } n} \underbrace{\frac{\text{SharesHeld}_{n,i,t-1}}{\text{Shares Outstanding}_{i,t-1}}}_{\equiv S_{n,i,t-1}} f_{n,t}. \tag{6}$$

$S_{n,i,t-1}$ is the proportion of all shares of stock i owned by mutual fund n in quarter $t-1$.

Variation in FIT demand shock size comes from stocks having heterogeneous exposures to fund flows. Since FIT is a cross-sectional shock, the variation comes from differences in fund ownership shares across stocks, not fund-level flows. Stocks with higher (lagged) ownership shares ($S_{n,i,t-1}$) by a given fund are more exposed to that fund's flows. Thus, stocks more exposed to high-flow funds have larger $\text{FIT}_{i,t}$.

FIT's exogeneity to cash flow news requires only that these ex-ante ownership shares do not correlate with cash flow news across stocks within each quarter (Chaudhry (2023); Chaudhary et al. (2023)).[6] This condition is plausible since FIT uses quarter $t-1$ ownership shares, which precede quarter t cash flow news. Thus, $\text{FIT}_{i,t}$ remains an uninformed demand shock even if flows contain cash flow information.

We construct FIT as in Lou (2012) using CRSP mutual fund flows, Thomson Reuters S12 holdings, and MFLINKS data from Russ Wermers.

All of our specifications include stock characteristic controls interacted with quarter fixed effects. These controls address the potential concern that within a quarter, stock characteristics may create a correlation between ownership shares and cash flow news. For example, good cash flow news about small stocks in quarter t may raise prices directly and also drive flows into small-cap funds. In this case the ownership shares are not exogenous: small stocks

[5]Following Li (2022), we do not multiply the numerator by a "partial scaling factor", as in Lou (2012), to adjust for differential flow-to-trade sensitivity for inflows and outflows. Li (2022) shows that applying this scaling factor or not does not materially impact inference on the price impact of FIT. In order to estimate price multipliers, we use shares outstanding in the denominator, so $\text{FIT}_{i,t} = 0.01$ represents the mutual fund sector buying 1% of stock i's shares.

[6]The sufficiency of cross-sectionally exogenous ownership shares follows from the result that exogenous shares are sufficient for shift-share instrument exogeneity (Goldsmith-Pinkham et al. (2020)).

are more exposed to both the good news shock and flows into small-cap funds. These correlated exposures to different "aggregate shocks" can threaten exogeneity. The solution is to control for the problematic stock characteristics (size in this example) interacted with time fixed effects to absorb the ownership share variation that correlates with exposures to aggregate cash flow news shocks (Chaudhry (2023)).

Order flow imbalance (OFI). Our third demand shock is the Lee-Ready signed order flow imbalance. Specifically, this measure takes all executed trades in the U.S. stock market, and uses the Lee and Ready (1991) algorithm to classify trades as buyer- or seller-initiated. In our implementation, we download daily Lee-Ready signed OFI from WRDS intraday indicators, aggregate to quarterly frequencies, and then normalize it by lagged shares outstanding. It is worth clarifying that OFI captures all trades that are executed *aggressively*, but it does not capture all trading flows. Many sophisticated institutional investors tend to execute trades slowly and passively to reduce price impact, and those flows will not be captured by OFI.

This OFI measure has been used heavily in the microstructure literature to study price effects at daily or higher frequencies. Recently, Li and Lin (2023) suggest that it can also be useful for studying demand effects at "asset pricing frequencies" (i.e. monthly or slower), as OFI does not revert and appears to create long-lasting price effects.

Relative to the other two measures, OFI has the benefit of having a lot more variation. This can be seen in the summary statistics of Table 1. One standard deviation is equal to 0.59% of shares outstanding for FIT but 4.15% for OFI. In the tails, the 1% and 99% percentiles indicate that over 2% of OFI realizations are even larger than 10% in absolute magnitude. This is particularly helpful for this paper because we are interested in price impacts of large demand shocks.

At the same time, a drawback of OFI is the concern about its information content. By its nature, OFI captures trading behavior by many investors, as opposed to FIT and BMI which zoom in on the trades by specific investors in specific circumstances that are plausibly unrelated to cash flow news. Li and Lin (2023) run 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 about the information content of OFI.

Overall, our three demand shocks have their own respective strengths and weaknesses. Successful instruments need to be exogenous and relevant. In our context, exogenity means not being correlated with cash flow-relevant news, and relevance refers to the amount of return variation it can explain. BMI is arguably the most exogenous, but index changes happen relatively rarely, so it has the lowest explanatory power over returns. Some researchers may be slightly more concerned about the exogeneity of FIT, but it has more variation as it happens to all stocks over all periods of time. Finally, OFI probably has the most amount of concern regarding its exogenenity, but it also has the highest amount of variation.

The findings we present in this paper are largely robust across these three measures, which we see as a strength: we do not rely on any single demand shock to draw our conclusions.

Institutional Holdings Data We use institutional holdings data from SEC Form 13F, 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. I allocate all remaining stock holdings to a residual "household" sector, which includes both direct stock holdings by households and those by non-13F institutions (i.e. institutions with less than $100 million AUM).

Other Data We download monthly stock returns and market capitalization from CRSP, and we aggregate monthly returns to quarterly frequency when needed. To control for factor-level effects, all regressions control for 15 commonly used stock characteristics from the website of Chen and Zimmermann (2022).[7] We also control for indicators for the Fama-French 12 industries; the industry classifications are downloaded from Professor Ken French's website.

Summary Statistics Table 1 reports summary statistics. Because we are interested in how the price multiplier varies by demand size, we show the percentile distributions on the right-most columns. OFI, reported as a fraction of lagged market capitalization, varies from -13.67% to 11.82% from the 1^{th} to the 99^{th} percentile. FIT and ΔBMI have similar ranges, with the former ranging from -1.32% to 2.13% and the latter from -2.1% to 2.32%.[8]

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.

[7]The stock characteristics we use include accruals, asset growth, beta, book-to-market, gross profitability, industry momentum, intermediate momentum, 1 year issuance, 5 year issuance, momentum, seasonal momentum, net operational assets, realized volatility, short-term reversal, and size. Following common practice in the literature on modeling the cross-section of returns, we transform each characteristic into uniform distributions over [-0.5, 0.5] in each cross-section (e.g. Kelly et al., 2019).

[8]Since ΔBMI measures the Russell index reconstitution-induced change in inelastic 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 one percentage point reconstitution-induced change in BMI raises total institutional ownership by 20 basis points. See Column 3 of Table 3 in Pavlova and Sikorskaya (2023) for justification for this scaling.

				Percentiles						
	Obs	Mean	StDev	1%	5%	25%	50%	75%	95%	99%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Return (%)	4,054	3.17	30.34	-63.28	-40.51	-12.17	1.38	14.87	50.59	108.54
Demand (Δ BMI, scaled)	472	0.05	0.71	-2.10	-1.14	-0.16	0.00	0.24	1.32	2.32
Demand (FIT, %)	4,054	0.08	0.59	-1.32	-0.72	-0.15	0.01	0.24	1.07	2.13
Demand (OFI, %)	4,054	-0.45	4.15	-13.67	-6.48	-1.68	-0.26	0.95	4.98	11.82
Market cap ($m)	4,054	4,049	25,126	8	20	98	367	1,592	14,857	67,060

Table 1. Summary statistics

Column (1) reports the average number of stocks per period. For FIT and OFI, one period is one quarter. For BMI, one period is the month of June in a specific year (which is when Russell index reconstitution occurs). The sample for all variables consists of quarterly data from 1993 to 2022, except the BMI sample which consists of data for each June from 1998 through 2018. For the demand shocks, ΔBMI refers to changes in benchmarking intensity in Pavlova and Sikorskaya (2023), FIT refers flow-induced trading in Lou (2012), and OFI refers to order flow imbalance in Li and Lin (2023). ΔBMI has been multiplied by 0.2 to have the same units as OFI and FIT, following Pavlova and Sikorskaya (2023).

4 Empirical Evidence of Concave Price Impact

In this section, we provide evidence of concave price impact: price multipliers shrink with demand shock size. Section 4.1 provides parametric evidence of this behavior using all three demand shocks discussed in Section 3. Section 4.2 exploits the large amount of variation in the OFI shock to provide evidence of this concavity from non-parametric sorts. Section 4.3 provides additional empirics to rule out alternative interpretations of these results.

4.1 Parametric Evidence

To test for nonlinear effects of demand shocks, we estimate the following cross-sectional regression of stock-level returns $r_{i,t}$ on demand shock $d_{i,t}$ (BMI, FIT, or OFI) and its interaction with shock magnitude $|d_{i,t}|$ in each time period t:

$$\forall t : r_{i,t} = b_{1,t} \cdot d_{i,t} + b_{2,t} \cdot d_{i,t} \times |d_{i,t}| + \boldsymbol{c}_t' \boldsymbol{x}_{i,t-1} + \tau_t + \epsilon_{i,t}. \tag{7}$$

The shock-size dependent price multiplier in (7) is $M = b_1 + b_2 \times |d_{i,t}|$. The key coefficient of interest is b_2. If price impacts are smaller for larger demand shocks, then b_2 is negative. Structurally, b_2 maps to the second derivative of the cost function $C(Q, P)$ in Section 2.

The controls $\boldsymbol{x}_{i,t-1}$ include the stock characteristics and industry indicators discussed in Section 3, as well as the BMI-specific controls discussed in Section 3 when using the BMI shock. Following previous work, we use a quarterly frequency for FIT and OFI, and a monthly frequency (just the month of June when Russell index reconstitution occurs) for BMI.

We estimate (7) in each cross section and report the average regression coefficients (in the style of Fama and MacBeth (1973)). Doing so ensures we use only cross-sectional variation to identify the regression coefficients, and so avoids the potential concern that omitted variables drive time variation in both price multipliers and demand shock size. In Appendix Table C.1, we find that estimating (7) using panel regressions with time fixed effects yield similar results.

	Dependent variable: stock return $r_{i,t}$		
$d =$	BMI	FIT	OFI
	(1)	(2)	(3)
$d_{i,t}$	1.68***	3.18***	3.00***
	(0.48)	(0.35)	(0.09)
$d_{i,t} \times \lvert d_{i,t} \rvert$	−58.73***	−43.73**	−13.92***
	(18.95)	(20.63)	(0.55)
Controls	Y	Y	Y
Obs	9,910	544,662	529,619
R^2	0.207	0.099	0.143
Marginal R^2 of demand	0.009	0.004	0.048

Table 2. Interacted price impact regressions
We estimate cross-sectional regressions (7) of stock returns on demand and demand interacted with its absolute value in each time period and report average coefficients pooled across all time periods. The regressions control for commonly used stock characteristics and Fama-French 12 industry indicators as described in Section 3. Column (1) reports results using the BMI demand shock and monthly returns. Columns (2) reports results using quarterly returns and the FIT and FOI demand shocks, respectively. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Table 2 reports the results of regression (7). The coefficient on $d_{i,t}$ (b_1) implies that the implied price multipliers for small demand shocks (i.e. $|d_{i,t}| \approx 0$) shocks lie in a range of 1.5 (for BMI) to 3 (for FIT and OFI). That is, for small X, buying $X\%$ of all shares outstanding of a stock raises price by around $1.5X\%$ to $3X\%$.

The coefficient on $d_{i,t} \times |d_{i,t}|$ (b_2) is negative, indicating that larger shocks are associated with smaller price multipliers. The coefficient magnitudes represent how many *basis points* the multiplier falls by when demand shock magnitude rises by 1% (i.e. $|d_{i,t}| = 1\%$ versus $|d_{i,t}| = 2\%$). This interaction coefficient is statistically significantly negative at the 5% level for FIT and at the 1% level for BMI and OFI.

The magnitude of b_2 is economically significant: assuming a constant price multiplier significantly overestimates the actual impact of large demand shocks. For example, the result in Column 1 (BMI) imply that buying 1% of shares outstanding has a multiplier of about $M = 1.1$ ($b_1 + b_2 \times 1 = 1.68 - 0.5873 \times 1 \approx 1.1$) and so raises price by about 1.1% ($M \times 1\%$), not 1.68% as a constant multiplier $M = b_1$ would suggest ($b_1 \times 1\% = 1.68\%$). A 2% shock has a multiplier of about $M = 0.5$ ($b_1 + b_2 \times 1 = 1.68 - 0.5873 \times 2 \approx 0.5\%$) and so raises price by about 1% ($M \times 2\% = 1\%$), not 3.36% as a constant multiplier $M = b_1$ would suggest ($b_1 \times 2\% = 1.68\%$). Thus, assuming a constant multiplier overestimates price impact by 70% for a 1% shock and by over 200% for a 2% shock.

The three demand shocks exhibit different degrees of nonlinearity. The b_2 results from Table 2 imply a 1% larger shock lowers the multiplier by about 59, 44, and 14 basis points for the BMI, FIT, and OFI shocks, respectively. Due to the sizeable standard errors for BMI and FIT, while the point estimates differ, the differences are not statistically significant.

Overall, all three demand shocks provide evidence that price multipliers shrink as demand shock size grows. In other words, price impact is concave in shock size. It is important to note, however, that even after accounting for this nonlinearity, multipliers are still far larger than suggested by classical asset pricing models which predict uninformed demand shocks have price multipliers on the order of 1/6000 (e.g. Petajisto, 2009). Therefore, our findings should be interpreted as sheding light on the endogenous variation of price multipliers, rather than arguing that multipliers are as small and demand is as elastic as predicted by classical asset pricing models.

4.2 Nonparametric Evidence

We next exploit the large degree of variation in the OFI demand shock to provide nonparametric evidence that price multipliers shrink as demand shock size grows.

We sort observations by realized OFI demand shocks $|d_{i,t}|$ into three ranges of $r \in \{[0, 2.5\%), [2.5\%, 5\%), [5\%, +\infty)\}$. We then estimate Fama-MacBeth regressions:

$$r_{i,t} = \sum_r M_{r,t} \cdot I_{|d_{i,t}| \in r} \cdot d_{i,t} + \boldsymbol{c}_t' \boldsymbol{x}_{i,t-1} + \tau_t + \epsilon_{i,t} \tag{8}$$

where the $\{M_r\}_r$ coefficients represent price multipliers for different demand shock sizes. The controls $\boldsymbol{x}_{i,t-1}$ include the stock characteristics and industry indicators discussed in Section 3. We report the estimated price multipliers in Panel A of Table 3.

The full sample results are reported in column (1). When $|d_{i,t}| < 2.5\%$, the estimated price multiplier is 3.58; when $|d_{i,t}|$ is between 2.5% and 5%, the associated multiplier declines to 2.55; when $|d_{i,t}| \geq 5\%$, the associated price multiplier is only 1.03.

To examine subsample robustness, in columns (2) through (5), we estimate the same regression over subperiods and find qualitatively similar results. In Panel B, we report pairwise price multiplier differences for each regression and compute standard errors using the Delta method. All differences are statistically significant at the 1% level. Overall, the results are consistent with the idea that price impact is concave: larger demand shocks have smaller price multipliers.

4.3 Ruling Out Alternative Interpretations

In this section, we conduct a series of robustness checks to rule out alternative interpretations of the empirical results in Sections 4.1 and 4.2.

Panel A: price impact regressions					
	Dependent variable: stock return $r_{i,t}$				
	Full sample	Sub-periods			
		1993-1999	2000-2007	2008-2015	2016-2022
	(1)	(2)	(3)	(4)	(5)
$M_{<2.5\%}$	3.58***	4.36***	4.06***	2.65***	3.30***
	(0.07)	(0.13)	(0.15)	(0.10)	(0.14)
$M_{[2.5\%,5\%)}$	2.55***	2.98***	2.58***	2.23***	2.44***
	(0.05)	(0.08)	(0.10)	(0.07)	(0.10)
$M_{\geq 5\%}$	1.03***	1.43***	0.97***	0.97***	0.78***
	(0.03)	(0.07)	(0.05)	(0.07)	(0.05)
Controls	Y	Y	Y	Y	Y
Obs	529,619	152,457	153,473	127,066	96,623
R^2	0.140	0.151	0.157	0.120	0.132
Panel B: Coefficient differences					
	(1)	(2)	(3)	(4)	(5)
$M_{[2.5\%,5\%)} - M_{<2.5\%}$	−1.03***	−1.38***	−1.48***	−0.42***	−0.86***
	(0.05)	(0.06)	(0.08)	(0.05)	(0.11)
$M_{\geq 5\%} - M_{[2.5\%,5\%)}$	−1.51***	−1.55***	−1.61***	−1.27***	−1.66***
	(0.03)	(0.04)	(0.06)	(0.07)	(0.08)
$M_{\geq 5\%} - M_{<2.5\%}$	−2.54***	−2.93***	−3.09***	−1.68***	−2.52***
	(0.06)	(0.08)	(0.12)	(0.08)	(0.14)

Table 3. Price multiplier by demand shock sizes

In panel A, we report Fama-MacBeth regressions which estimate (using quarterly data):

$$r_{i,t} = \sum_r M_{r,t} \cdot I_{|d_{i,t}| \in r} \cdot d_{i,t} + \boldsymbol{c}_t' \boldsymbol{x}_{i,t-1} + \tau_t + \epsilon_{i,t} \tag{10}$$

where $r_{i,t}$ is stock return and $d_{i,t}$ is the demand shock measured by order flow imbalance (OFI). The controls $\{x_{i,k,t}\}$ represent commonly used stock characteristics and Fama-French 12 industry indicators. OFI range r takes values in $[0, 2.5\%)$, $[2.5\%, 5\%)$, and $[5\%, \infty)$. Column (1) reports results based on the full sample, while columns (2) through (5) report results based on sub-samples. Panel B reports pairwise differences in price multipliers and the standard errors are computed using the Delta method. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

4.3.1 Alternative Interpretation: Larger Shocks are More Common in Times with Smaller Multipliers

One potential concern is that larger shocks are more common in time periods when price multipliers are generally smaller. Formally, if multipliers vary with an aggregate variable X_t:

$$r_{i,t} = b \cdot d_{i,t} + c \cdot X_t \cdot d_{i,t} + \boldsymbol{c}_t' \boldsymbol{x}_{i,t-1} + \tau_t + \epsilon_{i,t} \tag{11}$$

any correlation between X_t and shock size $|d_{i,t}|$ could falsely suggest multipliers shrink with shock size.

Our baseline specification (7) addresses this concern. Since we estimate (7) in each cross section, time-series correlation between X_t and shock size $|d_{i,t}|$ does not bias our b_2 estimates. Formally, the time-specific slope $b_{1,t}$ in (7) absorbs the $c \cdot X_t$ term in (11) ($b_{1,t} = b + c \cdot X_t$) and prevents it from biasing the b_2 estimates.

Thus, time-series correlations between aggregate variables and demand shock size do not explain the result that price multipliers decrease with shock size.

4.3.2 Alternative Interpretation: Larger Shocks are More Common for Stocks with Smaller Multipliers

Another potential concern is that larger shocks are more common for stocks that generally have smaller price multipliers. Formally, if multipliers vary with a stock-specific characteristic $X_{i,t}$:

$$r_{i,t} = b \cdot d_{i,t} + c \cdot d_{i,t} \cdot X_{i,t} + \boldsymbol{c}_t^{'} \boldsymbol{x}_{i,t-1} + \tau_t + \epsilon_{i,t} \tag{12}$$

any cross-sectional correlation between $X_{i,t}$ and shock size $|d_{i,t}|$ could falsely suggest multipliers shrink with shock size.

To address this concern, we augment our baseline specification (7) with interactions of the demand shock with common stock characteristics:

$$\forall t : r_{i,t} = b_{1,t} \cdot d_{i,t} + b_{2,t} \cdot d_{i,t} \times |d_{i,t}| + \boldsymbol{b}_{3,t}^{'} \cdot d_{i,t} \times \boldsymbol{x}_{i,t-1} + \boldsymbol{c}_t^{'} \boldsymbol{x}_{i,t-1} + \tau_t + \epsilon_{i,t}. \tag{13}$$

The characteristic interaction terms $\boldsymbol{b}_{3,t}^{'} \cdot d_{i,t} \times \boldsymbol{x}_{i,t-1}$ should absorb the $c \cdot d_{i,t} \cdot X_{i,t}$ term in (12) and so remove the potential omitted variable bias. As in Section 4.1, we estimate regression (13) in each cross section and report the average coefficients (in the style of Fama and MacBeth (1973)).

We include a variety of stock characteristics as controls: accruals, asset growth, beta, book-to-market, gross profitability, industry momentum, intermediate momentum, 1 year issuance, 5 year issuance, momentum, seasonal momentum, net operational assets, realized volatility, short-term reversal, and size.[9] The model in Section 2 implies that the other determinant of price multupliers besides the cost function $C(Q, P)$ is risk (i.e. return variance). These characteristics are often interpreted by the cross-sectional asset pricing literature as proxies for systematic risk, and so they may create cross-sectional variation in multipliers. We also control for indicators for the Fama-French 12 industries.

Figure 3 displays the estimated b_1 and b_2 coefficients from these regressions with interactions between the demand shock and stock characteristics. Each point represents the coefficient estimate from regression (13) when adding an additional characteristic (i.e. the right-most points represent coefficient estimates from the regression including all stock characteristics and industry indicators). We find evidence that price multipliers decrease with shock size across all specifications for all three demand shocks: $b_2 < 0$ for all specifications. Moreover, the coefficient estimates are quantitatively stable across specifications. Panel (a) demonstrates that for the BMI shock, b_1 ranges from 1.78 to 2.04 percent, while b_2 ranges from -60 to -25 basis points (increasing as characteristics are added). Panel (b) demonstrates that for the FIT shock, b_1 ranges from 3.08 to 3.38 percent, while b_2 ranges from -84 to -44 basis points (decreasing as characteristics are added). Panel (c) demonstrates that for the OFI shock, b_1 ranges from 2.97 to 3.01 percent, while b_2 ranges from -14 to -19 basis points (decreasing as characteristics are added).

Thus, cross-sectional correlations between stock characteristics and demand shock size does

[9] As discussed in Section 3, we transform each characteristic into uniform distributions over [-0.5, 0.5] in each cross-section.

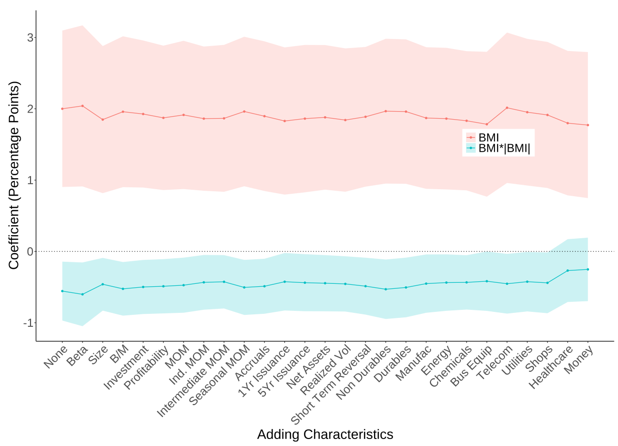
not appear to explain the result that price multipliers decrease with shock size.

4.3.3 Alternative Interpretation: Correlation with News Releases

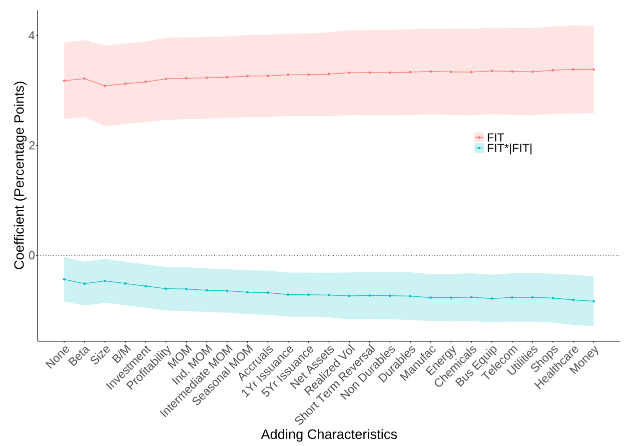
It is widely accepted that stock returns are often driven by news releases, which leads to the concern that perhaps our demand shocks *per se* do not impact prices, but rather are just correlate with news releases. Specifically, if demand is positively correlated with news-driven returns, and if the magnitude of demand happens to scale sublinearly with the magnitude of news-driven returns, then this can explain our finding. This concern proves less salient for BMI and FIT which exploit variation in trading by specific investors in specific circumstances that are plausibly unrelated to cash flow news. However, this proves potentially more relevant for OFI, which captures trading behavior by many investors over a whole quarter.

To assess this concern, we examine whether we observe concave price impact on days with little news. We use three measures of news releases at the daily frequency. The first is whether there is an earnings release. The second is the number of analyst updates from IBES, and the third is the number of media reports in Ravenpack. For the latter two, we convert them into quintile indicators after sorting within each quarter-stock to account for the fact that larger stocks and later periods tend to have more updates. This conversion makes the news measures more comparable across stocks and over time.

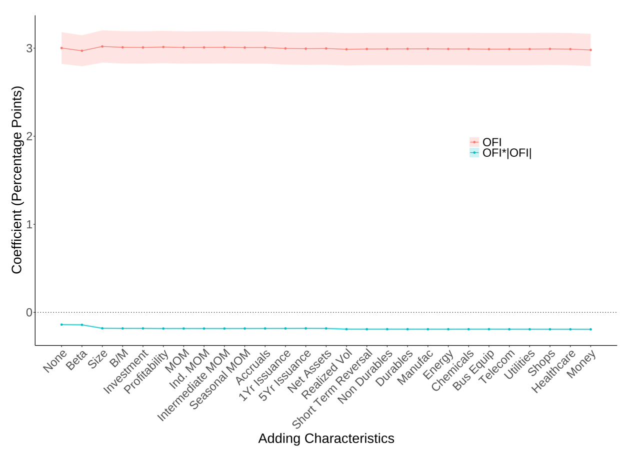
We first verify that these measures do capture price-relevant news. In Table 4, we regress the absolute value of daily stock returns on news indicators. As a refernece point, the dependent variable's average value is 2.2%. In the last column of Table 4 which contains all news indicators, we find that earnings days are associated with 0.92% higher return absolute values. For the other two news measures, days on which they are higher — and mostly in the top quintile — are also associated with higher return variation. On days when IBES and Ravenpack are in the top quintile of releases, absolute returns are higher by 0.46% and 0.51%, respectively.



(a) BMI



(b) FIT



(c) OFI

Figure 3. Panels (a), (b), and (c) plot the results of regression (13) for the BMI, FIT, and OFI demand shocks, respectively. The coefficient on the demand shock is plotted in red, while the coefficient on the demand shock interacted with its magnitude is plotted in blue. The left-most points represent the coefficient estimates when no interactions between the demand shock and any stock characteristics are included. Each subsequent point represents the estimates from regression (13) when adding the interaction with characteristic labelled on the x-axis (i.e. the right-most points represent coefficient esimtates from the regression including all stock characteristics and industry indicators). Both coefficients are expressed in terms of percentage points (i.e. b_2 has been divided by 100 to convert it from basis points to percentage points, so its units match those of b_1).

	Dependent variable: $\lvert r_{i,t}\rvert(in\%)$			
	(1)	(2)	(3)	(4)
Earnings	1.46***			0.92***
	(0.04)			(0.04)
IBES bin 2		−0.00		−0.00
		(0.00)		(0.00)
IBES bin 3		−0.00**		−0.00**
		(0.00)		(0.00)
IBES bin 4		0.01***		0.01***
		(0.00)		(0.00)
IBES bin 5		0.46***		0.36***
		(0.01)		(0.01)
Ravenpack bin 2			0.00	0.00
			(0.00)	(0.00)
Ravenpack bin 3			0.01**	0.01**
			(0.00)	(0.00)
Ravenpack bin 4			0.04***	0.04***
			(0.01)	(0.01)
Ravenpack bin 5			0.63***	0.51***
			(0.02)	(0.02)
Controls	Y	Y	Y	Y
Time-stock FE	Y	Y	Y	Y
Obs	16,155,699	16,155,699	16,155,699	16,155,699
Within R^2	0.005	0.005	0.009	0.014

Table 4. Information measures and return variability

We use daily panel regressions to estimate the relationship between the absolute value of stock returns and indicators of news releases. All regressions control for quarter-stock fixed effects, as well as the list of controls in Table 2. Standard errors are clustered by quarter and stock. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

We now re-estimate the results for FIT and OFI in Table 2 but remove days with more news from our computation of quarterly returns. The results are reported in Table 5. Columns (1) and (5) report results for FIT and OFI, respectively, when no filters are applied, and the results indicate that price impacts are concave.[10] In columns (2) and (6), we use quarterly returns computed only during days without earnings. In columns (3) and (7), we further remove days where IBES or Ravenpack news measures are above 80% percentiles, and columns (4) and (8) further remove days where either news measure is above 60% percentiles. Overall, the coefficient estimates are quantitatively similar across specifications, thereby indicating that removing days with news releases does not impact our conclusion that price multipliers

[10]The results differ slightly from Table 2 because we merge our sample with news measures, which results in a smaller dataset.

decrease in demand shock size.

	Dependent variable: stock return $r_{i,t}$							
	$d_{i,t}$ = FIT				$d_{i,t}$ = OFI			
Filters:	none	No earnings	news $\leq$ 80%	news $\leq$ 60%	none	No earnings	news $\leq$ 80%	news $\leq$ 60%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$d_{i,t}$	3.34***	3.41***	4.10***	3.70***	2.23***	2.27***	2.28***	2.25***
	(0.38)	(0.38)	(0.48)	(0.61)	(0.09)	(0.09)	(0.11)	(0.12)
$d_{i,t} \times \lvert d_{i,t} \rvert$	−65.39***	−65.87***	−66.29*	−69.81	−9.83***	−9.99***	−8.79***	−9.42***
	(23.15)	(22.52)	(38.92)	(48.37)	(0.72)	(0.72)	(0.94)	(1.06)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Obs	228,379	228,379	228,379	228,379	229,427	229,427	229,427	229,427
R^2	0.129	0.129	0.093	0.055	0.160	0.159	0.111	0.064
Marg $R^2(d_{i,t})$	0.006	0.006	0.004	0.002	0.036	0.035	0.021	0.011

Table 5. Interacted price impact regressions in periods with less news
We re-estimate the Fama-MacBeth regressions in Table 2 for FIT and OFI, but only use returns on days with less news. Columns (1) through (4) study FIT while columns (5) through (8) study OFI. Columns (1) and (5) do not apply filters. Columns (2) and (6) filter out earnings days when computing quarterly returns. Columns (3) and (7) further filter out days where IBES or Ravenpack news indicators are above the 80% percentile; columns (4) and (8) further filter out those above the 60% percentile. As explain in the text, regression coefficients and standard errors are adjusted to be comparable. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

It is worth explaining that, as we remove some days in computing quarterly returns, this leads to a mechanical downward bias in the estimated coefficients. The reported coefficients and standard errors in Table 5 have adjusted for this bias.[11]

5 A Nonlinear Asset Demand System

In this section we propose a nonlinear asset demand system with endogenous price elasticities of demand to study the quantitative implications of the concave price impact we document.

[11] To see this, suppose the true relationship at the daily frequency is $r_{i,t} = M \cdot d_{i,t} + \epsilon_{i,t}$ where $r_{i,t}$ is the daily log return, $d_{i,t}$ is the demand shock, M is the price muliptlier, and $\epsilon_{i,t}$ is a residual term. For simplicity, assume all components are i.i.d. Let $D_i = \sum_{t=1}^{T} d_{i,t}$ be the quarterly demand and $R_i = \sum_{t \in \mathcal{T}} r_{i,t}$ be the quarterly log return when only using a subset of days, $\mathcal{T} \subset \{1, ..., T\}$. Then, the regression coefficient at the quarterly frequency is:

$$\frac{Cov(D_i, R_i)}{Var(D_i)} = \frac{|\mathcal{T}|}{T} \cdot M$$

which is downward biased by an adjustment factor of $\frac{|\mathcal{T}|}{T}$. To fix this, we multiply the raw regression coefficients and standard errors by $\frac{T}{\mathcal{T}|}$ before reporting them in Table 5.

5.1 Model

To derive a tractable asset demand system, we extend the model from Section 2 to feature multiple assets and choose a specific form for the cost function $C(Q, P)$.

Agents and Prefferences There are $i = 1, \ldots, I$ investors and $n = 1, \ldots, N$ assets. Each investor i chooses a vector of quantities for each asset $\mathbf{Q}_i$ to solve

$$\max_{\mathbf{Q}_i} \mathbf{Q}_i' \mathbb{E}_i \left[\mathbf{D} - \mathbf{P} \right] - \frac{\gamma_i}{2} \mathbf{Q}_i' \mathbb{V}_i \left[\mathbf{D} - \mathbf{P} \right] \mathbf{Q}_i - \pi_i \left\| \mathbf{Q}_i - \mathbf{B}_i \right\|_{\alpha_i}^{\alpha_i} \tag{14}$$

where $\mathbf{D}$ and $\mathbf{P}$ are the vectors of dividends and prices for each asset. The i subscript indicates the expectation and variance are taken under investor i's subjective beliefs.

The cost function is the scaled L^α norm of the deviation of the chosen quantities $\mathbf{Q}_i$ from some refference portfolio $\mathbf{B}_i$:

$$C(\mathbf{Q}_i, \mathbf{P}) = \pi_i \left\| \mathbf{Q}_i - \mathbf{B}_i \right\|_{\alpha_i}^{\alpha_i} = \sum_n \left| Q_{i,n} - B_{i,n} \right|^{\alpha_i}. \tag{15}$$

In general, the refference portfolio can be any benchmark or anchor the investor does not want to deviate from. Empirically, we use the investors' portfolio from the previous quarter as the refference portfolio, which gives $C(\mathbf{Q}_i, \mathbf{P})$ the interpretation of an adjustment cost.

$\pi_i \geq 0$ controls the strength of this penalty. α determines if the marginal cost curve is convex (for $\alpha_i > 2$) or concave (for $\alpha_i < 2$) in $|\mathbf{Q}_i - \mathbf{B}_i|$. Thus, α_i determines if this investor becomes less (for $\alpha_i > 2$) or more (for $\alpha_i < 2$) elastic for larger deviations from the refference portfolio $|\mathbf{Q}_i - \mathbf{B}_i|$ (as dicussed in Section 2.2). We assume $\alpha > 1$ so that marginal cost is increasing and demand slopes downward.

Beliefs Following Koijen et al. (2024), we assume each investor i's beliefs about future cash flows follow a factor structure

$$\mathbf{D} = \boldsymbol{\mu}_i + \boldsymbol{\rho}_i F + \boldsymbol{\eta} \tag{16}$$

where the exepcted cash flow $\boldsymbol{\mu}_i$ and factor loading $\boldsymbol{\rho}_i$ are linear in stock characteristics $\mathbf{x}_n$:

$$\mu_{i,n} = \mathbf{\Phi}'_{\mu,i}\mathbf{x}_n + \phi^{\mu}_{i,n}$$
$$\rho_{i,n} = \mathbf{\Phi}'_{\rho,i}\mathbf{x}_n + \phi^{\rho}_{i,n}.$$

$\boldsymbol{\eta}$ is an idiosyncratic shock (uncorrelated with F) with mean zero and covariance matrix $\sigma_i^2\mathbf{I}$.

Optimal Portfolio Choice Taking the first-order condition of (14) yields the following characteristics-based demand function for investor i (derived in Appendix D.1):

$$Q_{i,n} = \underbrace{\beta_{0,i}}_{\equiv -\frac{1}{\gamma\sigma^2}} P_n + \underbrace{\boldsymbol{\beta}'_{1,i}}_{\equiv \frac{\mathbf{\Phi}'_{\mu,i} - \gamma\boldsymbol{\rho}'\boldsymbol{Q}_i\mathbf{\Phi}'_{\rho,i}}{\gamma\sigma^2}} \mathbf{x}_n - \underbrace{\lambda_i}_{\equiv \frac{\pi\alpha}{\gamma\sigma^2}} |Q_{i,n} - B_{i,n}|^{(\alpha_i - 1)} \operatorname{sign}(Q_{i,n} - B_{i,n}) + \underbrace{\epsilon_{i,n}}_{\equiv \frac{\phi'_{\mu,i} - \gamma\boldsymbol{\rho}'\boldsymbol{Q}_i\phi_{\rho,i}}{\gamma\sigma^2}}. \tag{17}$$

To illustrate how the price elasticity of demand depends on α, Figure 4 plots the derivative of quantity with respect to price (expressed as a positive number)

$$-\frac{dQ_{i,n}}{dP_n} = -\frac{\beta_{0,i}}{1 + \lambda_i(\alpha_i - 1)\,|Q_{i,n} - B_{i,n}|^{(\alpha_i - 2)}} \tag{18}$$

as a function of the deviaton from the refference portfolio $|Q_{i,n} - B_{i,n}|$. For $\alpha_i > 2$, demand becomes *less* elastic for larger deviations from $B_{i,n}$ (consistent with *convex* price impact). For $\alpha < 2$, demand becomes *more* elastic for larger deviations (consistent with *concave* price impact). For $\alpha = 2$, elasticity is constant (consistent with *linear* price impact).

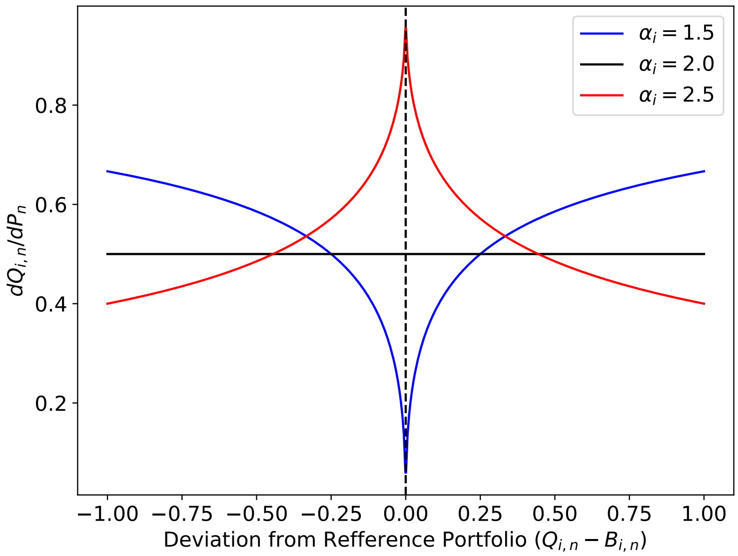


Figure 4. Plot of sensitivity of demand to price (18) for $\lambda_i = 1, \beta_{0,i} = -1$, and different values of α_i.

5.2 Empirical Specification

Based on (17), we model the portfolio weight of investor i for stock $n = 1, \ldots, N$ in quarter t as follows to match the lognormal distribution of weights in institutional holdings data (as in Koijen and Yogo (2019); Koijen et al. (2024))

$$w_{i,n,t} = \frac{\delta_{i,n,t}}{1 + \sum_m^N \delta_{i,m,t}}, \tag{19}$$

where

$$\delta_{i,n,t} = \exp\left[\beta_{0,i,t} me_{n,t} + \boldsymbol{\beta}'_{1,i,t}\mathbf{x}_{n,t} - |K_{i,t} + \Delta \log \delta_{i,n,t} - \Delta me_{n,t}|^{\alpha_{i,t}-1} + C_{i,t}\right] \epsilon_{i,n,t}. \tag{20}$$

There is also an outside asset (consisting of stocks that are foreign, real estate investment trusts, or have missing characteristics or returns, as in Koijen and Yogo (2019)), with weight $w_{i,0,t} = 1 - \sum_{n=1}^N w_{i,n,t}$ so that portfolio weights sum to one.

$me_{n,t}$ is log market equity and $\mathbf{x}_{n,t}$ contains the stock characteristics used in Koijen and Yogo (2019): log book equity, profitability, investment, market beta, and the dividend-to-book equity ratio.

We model the penalty term as an adjustment cost from this investor's portfolio in the previous quarter. In particular, we assume it costly to change the number of shares an investor holds:

$$\begin{aligned} Q_{i,n,t} &= AUM_{i,t} w_{i,n,t} / P_{i,n,t} \\ \leftrightarrow \Delta \log Q_{i,n,t} &= \Delta \log AUM_{i,t} + \Delta \log w_{i,n,t} - \Delta \log P_{n,t} \\ &= \underbrace{K_{i,t}}_{\equiv \Delta \log AUM_{i,t} - \Delta \log\left(1+\sum_m^N \delta_{i,m,t}\right)} + \Delta \log \delta_{i,n,t} - \underbrace{\Delta me_{n,t}}_{\equiv \Delta P_{n,t}} . \end{aligned} \tag{21}$$

$AUM_{i,t}$ is assets under management in dollars. The second line follows from (19) and noting that $\Delta \log P_{n,t} = \Delta me_{n,t}$ (after adjusting for stock splits). Thus, we assume there is no cost to passive weight changes driven by price fluctuations that are not accompanied by changes in number of shares held.

Identification The stock characteristics $\mathbf{x}_{n,t}$ are assumed exogenous to latent demand $\epsilon_{i,n,t}$. However, market equity is endogenous to latent demand through market clearing. Thus, we use the market equity instrument of Koijen and Yogo (2019):

$$\widehat{\mathrm{me}}_{i,n,t} = \log\left(\sum_{j \neq i} AUM_{j,t-1} \frac{1_{j,n,t}}{1 + \sum_{m=1}^N 1_{j,m,t}}\right),$$

where $1_{j,n,t}$ is an indicator for if stock n is the investment universe of investor j and $AUM_{j,t-1}$ is the (lagged one quarter) assets under management of investor j. One can interpret this instrument as the counterfactual market equity of stock n if all investors held an equal-weighted portfolio of the stocks in their investment universe. This instrument exploits only the (lagged) wealth distribution and the investment universes of other investors, both of which we take as exogenous. This assumption proves reasonable because investment universes are defined by investment mandates, which are predetermined rules that don't change in response to current latent demand shocks. Thus, if stock n exogenously falls into the investment universe of more or larger investors, it will face greater demand and will have greater market equity. Following Koijen and Yogo (2019), we measure the investment universe of investor i as the set of all stocks this investor currently holds or has ever held in the previous eleven quarters.

Estimation We estimate (20) in the cross section of each investor i's holdings in each quarter t using generalized method of moments (GMM). For K stock characteristics, there are $K+3$ parameters to estimate: $C_{i,t}, \alpha_{i,t}, \beta_{0,i,t}, \boldsymbol{\beta}_{1,i,t} \in \mathbb{R}^K$. [12] Thus, we use the following $K+3$ moment conditions:

$$\mathbb{E}\left[\epsilon_{i,n,t} \mid \widehat{1, \mathrm{me}}_{i,n,t}, \widehat{\mathrm{me}}_{i,n,t-1}, \mathbf{x}_{n,t}\right] = 1.$$

We use the lagged market equity instrument $\widehat{\mathrm{me}}_{i,n,t-1}$ because lagged market equity appears in (20) via the adjustment cost term.

Following Koijen and Yogo (2019), we estimate (20) separately for each investor with at least 1,000 strictly positive holdings in the current quarter t. We pool investors with fewer than 1,000 holdings into groups based on quantiles of assets under management conditional on investor type (banks, insurance companies, investment advisors, mutual funds, pension funds, other 13F institutions, or the residual household sector). The number of groups at

[12] We measure $K_{i,t}$ directly as in (21). In particular, we use the Davis and Haltiwanger (1992) method to calculate approximate log changes for $K_{i,t}$ and $\Delta \log \delta_{i,n,t}$:

$$K_{i,t} \equiv \Delta \log AUM_{i,t} - \Delta \log \left(1 + \sum_m^N \delta_{i,m,t}\right) \approx \frac{AUM_{i,t}/\left(1+\sum_m^N \delta_{i,m,t}\right) - AUM_{i,t-1}/\left(1+\sum_m^N \delta_{i,m,t-t}\right)}{\frac{1}{2}\left(AUM_{i,t}\left(1+\sum_m^N \delta_{i,m,t}\right) + AUM_{i,t-1}\left(1+\sum_m^N \delta_{i,m,t-t}\right)\right)}$$

$$\Delta \log \delta_{i,n,t} = \frac{\delta_{i,n,t} - \delta_{i,n,t-1}}{\frac{1}{2}\left(\delta_{i,n,t} + \delta_{i,n,t-1}\right)}.$$

The Davis and Haltiwanger (1992) method is the first-order approximation to the log difference and allows a uniform treatment of both cases where $\delta_{i,n,t-1} \neq 0$ and where $\delta_{i,n,t-1} = 0$.

each date is set to target an average of 2,000 holdings per group.

5.3 Estimation Results

Figure 5 displays the pooled results from estimating the nonlinear asset demand system (20) across all investors and quarters. Note that $\alpha_{i,t} < 2$ for the vast majority of investors and quarters, which is consistent with investors becoming more price-elastic when making larger adjsutments. Thus, this result is consistent with the concave price impact we document in reduced-form in Section 4.

5.4 Quantifying the Importance of Concave Price Impact

To quantify the importance of the concave price impact we document, we use the estimated demand system to conduct a simple counterfactual analysis.

We start with the equilibrium prices and holdings at the end of quarter t. Then, with stock characteristics and prices held fixed, at the start of quarter $t+1$ we shock all investors' latent demand by an amount that would induce each investor to buy (or sell) $X\%$ of a particular stock n (as a proportion of their end of quarter t holdings). We then calculate the new equilibrium prices (for all stocks) that clear the market after this shock (using the algorithm in Appendix D.2)[13]:

$$\forall m : ME_{m,\text{Post Shock}} = \sum_i A_{i,t} w_{i,m,\text{Post Shock}}.$$

Given the counterfactual prices, we compute the price multiplier for this shock for stock n:

$$M_{n,X} = \frac{\frac{ME_{n,\text{Post Shock}} - ME_{n,\text{Pre Shock}}}{ME_{n,\text{Pre Shock}}}}{X},$$

where $ME_{n,\text{Pre Shock}}$ is the observed market equity at the end of quarter t.

[13]Following Koijen and Yogo (2019), we hold wealth fixed when computing counterfactual prices.

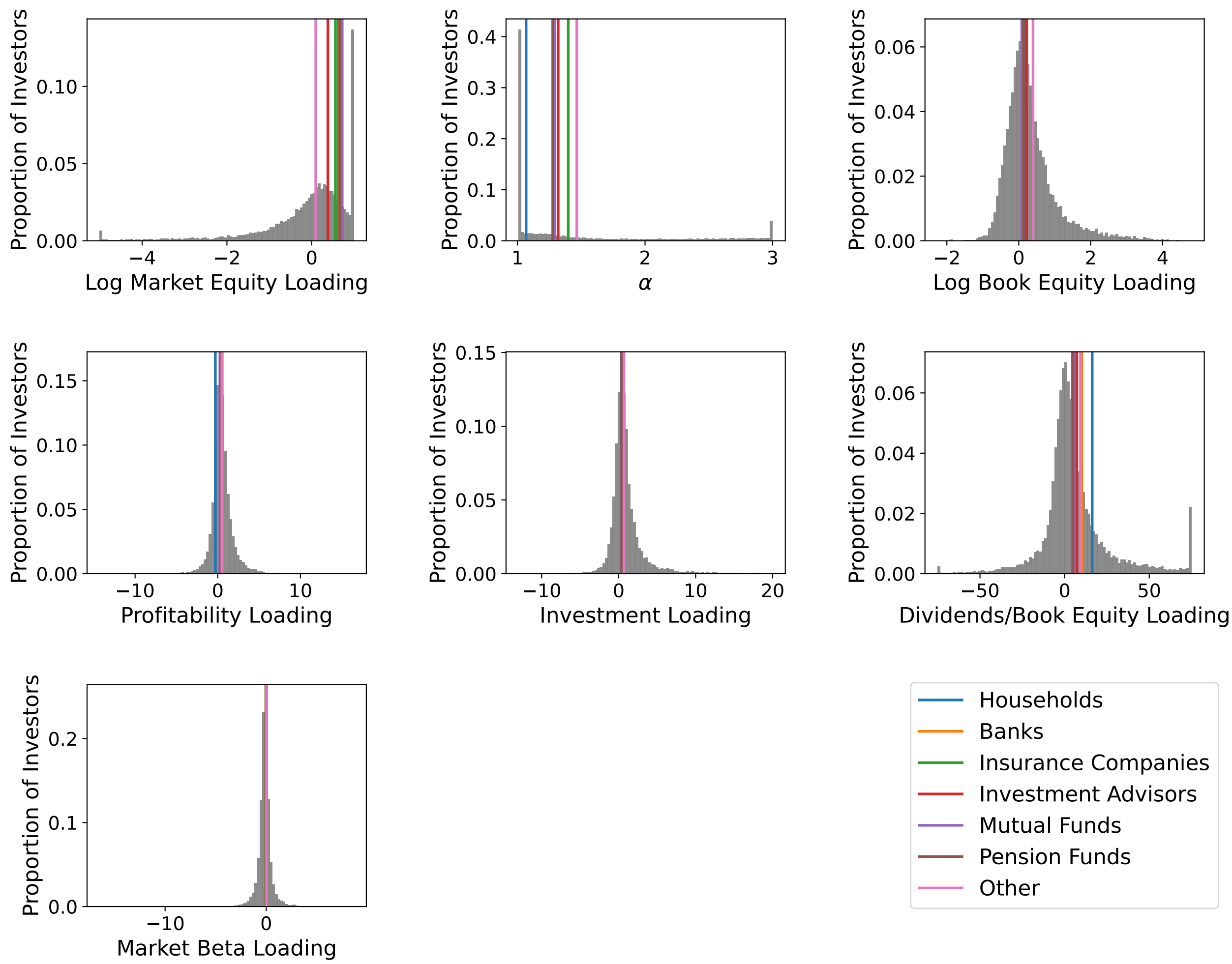


Figure 5. Pooled estimation results across all investors and quarter for nonlinear asset demand system (20). Ther vertical lines indicate AUM-weighted averages for each parameter and investor type. These averages are calculated by first calcualting the AUM-weighted average parameter value within each investor type and quarter, and then averaging across all quarters.

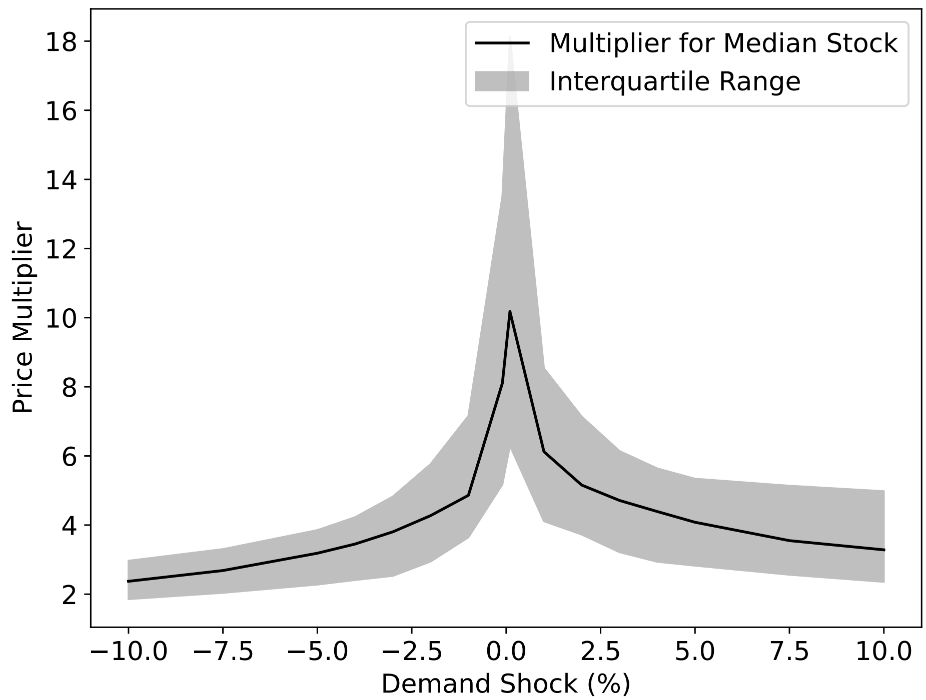


Figure 6. Price multipliers for different demand shock sizes calculated using the estimated asset demand system. The solid black line displays the median multiplier across all stocks. The shaded area represents the interquartile range for multipliers across shocks (i.e. the lower and upper bounds are the $25th$ and $75th$ percentiles for multipliers across all stocks).

We compute $M_{n,X}$ for all stocks n in the second quarter of 2015 and shock shocks X between -10% and 10%.[14]

Figure 6 dispalys the median multiplier $M_{n,X}$ across all stocks n for each shock size. Small shock ($X = -10$ or 10 basis points) multipliers are roughly three times larger than large shock ($X = -10$ or 10 percentage points) multipliers ($M \approx 10$ and $M \approx 3$, respectively).

6 Conclusion

In this paper, we document a new stylized fact about inelastic demand in financial markets: larger uninformed demand shocks have smaller price multipliers. That is, we find the price impact of uninformed demand shocks is concave in shock size.

[14]We are working optimize the numerical calculation of the coutnerfactual prices in order to extend this analysis to all quarters.

We provide evidence of this concavity using three cash flow-unrelated demand shocks from previous work: index-reconstitution induced changes in benchmarking intensity, mutual fund flow-induced trading, and order flow imbalance. While all three demand shocks display large price impact, the per-unit price impact is smaller for shocks of larger magnitude.

Motivated by these findings, we propose a nonlinear asset demand system. We find the concavity we document is quantitatively important for counterfactual analyses in financial markets: extrapolating local price multiplier estimates may overstate the impact of large quantity shifts on prices.

This concavity provides a new moment to discipline asset pricing and macro-finance models, many of which rely on frictions that generate convex or linear price impact. We establish general conditions under which concave price impact can arise. Two broad classes of mechanisms satisfy these conditions: those in which it is less costly to take larger positions than smaller ones (e.g. fixed costs), and those in which investors can pay to expand their risk-bearing capacity when presented with profitable opportunities (e.g. slow-moving capital or costly information acquisition).

Overall, our findings illuminate an important dimension of endogenous variation in price multipliers. Furthermore, we provide new empirical methods to study the price impact of large demand shocks while accounting for this endogenous variation.

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.

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.

Asquith, P., and D. W. Mullins Jr. 1986. Equity issues and offering dilution. *Journal of financial economics* 15:61–89.

Bacchetta, P., S. Tieche, and E. Van Wincoop. 2023. International portfolio choice with frictions: Evidence from mutual funds. *The Review of Financial Studies* 36:4233–4270.

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.

Ben-David, I., J. Li, A. Rossi, and Y. Song. 2022. Ratings-driven demand and systematic price fluctuations. *The Review of Financial Studies* 35:2790–2838.

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

Brunnermeier, M. K., and Y. Sannikov. 2014. A macroeconomic model with a financial sector. *American Economic Review* 104:379–421.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng. 2024. On binscatter. *American Economic Review* 114:1488–1514.

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 J. Li. 2023. Corporate bond multipliers: Substitutes matter. *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* .

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.

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

DellaVigna, S., and J. M. Pollet. 2009. Investor inattention and Friday earnings announcements. *The journal of finance* 64:709–749.

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., and B. Strulovici. 2012. Capital mobility and asset pricing. *Econometrica* 80:2469–2509.

Eckbo, B. E., R. W. Masulis, and Ø. Norli. 2007. Security offerings. *Handbook of empirical corporate finance* pp. 233–373.

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* .

Frazzini, A., and O. A. Lamont. 2008. Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of financial economics* 88:299–322.

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

Gârleanu, N., and L. H. Pedersen. 2013. Dynamic trading with predictable returns and transaction costs. *The Journal of Finance* 68:2309–2340.

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.

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

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, B. 2018. Dynamic information acquisition and asset prices. *Unpublished Working Paper, University of Maryland* .

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., and A. Krishnamurthy. 2013. Intermediary asset pricing. *American Economic Review* 103:732–770.

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

Ikenberry, D., J. Lakonishok, and T. Vermaelen. 1995. Market underreaction to open market share repurchases. *Journal of financial economics* 39:181–208.

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.

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. 2023. Prices are less elastic at more aggregate levels. *Working paper* .

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.

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.

Sikorskaya, T. 2023. Institutional investors, securities lending and short-selling constraints. *Securities Lending, and Short-Selling Constraints (November 8, 2023)* .

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 der Beck, P. 2021. Flow-driven ESG returns. *Swiss Finance Institute Research Paper* .

Van der Beck, P. 2022. On the estimation of demand-based asset pricing models. *Swiss Finance Institute Research 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.

APPENDIX

A Additional Theoretical Results

A.1 Example of $\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) > 0$: Convex Adjustment Costs

Consider the following cost function:

$$C(Q, P) = \lambda(Q)^{2\alpha},$$

for some constants $\lambda > 0$ and $\alpha > 1$.

In this case, we have marginal cost

$$MC(Q, P) \equiv \frac{\partial}{\partial Q}C(Q, P) = 2\alpha\lambda(Q)^{2\alpha-1},$$

and so equilibrium price

$$P = \mathbb{E}\left[\tilde{D}\right] - \Theta \cdot \gamma\mathbb{V}\left[\tilde{D}\right] - 2\alpha\lambda(\Theta)^{2\alpha-1}.$$

The price multiplier is thus:

$$\begin{aligned} M &= -\frac{\partial}{\partial\Theta}P = \gamma\mathbb{V}\left[\tilde{D}\right] + 2\alpha(\alpha-1)\lambda(\Theta)^{2(\alpha-1)} \\ &= \frac{\partial}{\partial\Theta}P = \gamma\mathbb{V}\left[\tilde{D}\right] + 2\alpha(\alpha-1)\lambda|\Theta|^{2(\alpha-1)}. \end{aligned}$$

Hence, M is increasing in supply shock size $|\Theta|$:

$$\frac{\partial}{\partial|\Theta|}M = 2\alpha(\alpha-1)(2\alpha-2)\lambda|\Theta|^{2\alpha-3} > 0$$

because $\alpha > 1$ by assumption.

Note that this cost function satisfies $\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) > 0$:

$$\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta, P) = 2\alpha(\alpha-1)(2\alpha-2)\lambda|\Theta|^{2\alpha-3} > 0.$$

A.2 Example of $\frac{\partial^2}{\partial|\Theta|\partial P} MC(\Theta, P) < 0$: Leverage Constraints

Assume the representative investor is subject to a "leverage constraint" — a limit on dollar position size:

$$
\begin{aligned}
&\max_Q \mathbb{E}\left[Q\left(\tilde{D}-P\right)\right] - \frac{\gamma}{2}\mathbb{V}\left[Q\left(\tilde{D}-P\right)\right] \\
&s.t. QP \leq \alpha
\end{aligned}
$$

If the constraint does not bind ($QP < \alpha$), the investor demands

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

and so

$$P = \mathbb{E}\left[\tilde{D}\right] - \Theta\gamma\mathbb{V}\left[\tilde{D}\right].$$

If the constraint does bind ($QP \geq \alpha$), the investor demands

$$Q = \frac{\alpha}{P}.$$

and so

$$P = \frac{\alpha}{\Theta}.$$

The unconstrained equilibrium prevails when (by market clearing $Q = \Theta$)

$$
\begin{aligned}
\alpha &> \Theta P \\
&= \Theta\mathbb{E}\left[\tilde{D}\right] - \Theta^2\gamma\mathbb{V}\left[\tilde{D}\right] \\
\leftrightarrow \Theta &< \Theta_- \text{ or } \Theta > \Theta_+,
\end{aligned}
$$

where

$$\Theta_-, \Theta_+ = \frac{-\mathbb{E}\left[\tilde{D}\right] \pm \sqrt{\left(\mathbb{E}\left[\tilde{D}\right]\right)^2 - 4\alpha\gamma\mathbb{V}\left[\tilde{D}\right]}}{-2\gamma\mathbb{V}\left[\tilde{D}\right]},$$

and $0 < \Theta_- < \Theta_+$.

So the price multiplier is

$$M = -\frac{\partial}{\partial \Theta} P = \begin{cases} \gamma \mathbb{V}\left[\tilde{D}\right] & , \Theta < \Theta_- \text{ or } \Theta > \Theta_+ \\ \frac{\alpha}{\Theta^2} & , \Theta \in [\Theta_-, \Theta_+] \end{cases}$$

Thus, for sufficiently large α

$$\alpha > \Theta_-^2 \gamma \mathbb{V}\left[\tilde{D}\right],$$

M increases at the point where the constraint binds ($\Theta = \Theta_-$) and so

$$\frac{\partial}{\partial |\Theta|} M > 0.$$

We can reexpress this setup in the cost function notation of Section 2 as

$$\max_Q \mathbb{E}\left[Q\left(\tilde{D} - P\right)\right] - \frac{\gamma}{2} \mathbb{V}\left[Q\left(\tilde{D} - P\right)\right] - C(Q, P),$$

where

$$C(Q, P) = \lambda(Q, P)\left(\alpha - QP\right)$$
$$\lambda(Q, P) = \begin{cases} 0, & \Theta < \Theta_- \text{ or } \Theta > \Theta_+ \\ \frac{\mathbb{E}[\tilde{D}] - \frac{\alpha}{P}\gamma \mathbb{V}[\tilde{D}]}{P} - 1, & \Theta \in [\Theta_-, \Theta_+] \end{cases}$$

$\lambda(Q, P)$ represents the shadow cost of relaxing the leverage constraint.

Hence, marginal cost is

$$MC(Q, P) \equiv \frac{\partial}{\partial Q} C(Q, P) = \begin{cases} 0, & \Theta < \Theta_- \text{ or } \Theta > \Theta_+ \\ \mathbb{E}\left[\tilde{D}\right] - P + \frac{\alpha}{P}\gamma \mathbb{V}\left[\tilde{D}\right] - 2\gamma \mathbb{V}\left[\tilde{D}\right] Q, & \Theta \in [\Theta_-, \Theta_+]. \end{cases}$$

and so we have

$$\frac{\partial}{\partial P} MC(Q, P) = \begin{cases} 0, & \Theta < \Theta_- \text{ or } \Theta > \Theta_+ \\ -1 - \frac{\alpha}{P^2}\gamma \mathbb{V}\left[\tilde{D}\right], & \Theta \in [\Theta_-, \Theta_+]. \end{cases}$$

Thus, as Θ increases from below to $\Theta-$ and the constraint binds (note this means $|\Theta|$ increases since $\Theta- > 0$), $\frac{\partial}{\partial P} MC(Q,P)$ falls from 0 to less than 0. Therefore,

$$\frac{\partial^2}{\partial|\Theta|\partial P} MC(\Theta, P) < 0$$

at the point where the constraint binds.

A.3 Example of $\frac{\partial^2}{\partial|\Theta|\partial\Theta} MC(\Theta, P) = \frac{\partial^2}{\partial|\Theta|\partial P} MC(\Theta, P) = 0$: Benchmarking

Let the investor's compensation be linear combination of his absolute return and his return relative to a benchmark (as in Pavlova and Sikorskaya (2023))

$$c(Q) = \underbrace{aQ\left(\tilde{D} - P\right)}_{\text{Absolute Return}} + \underbrace{b\left[Q\left(\tilde{D} - P\right) - \bar{Q}\left(\tilde{D} - P\right)\right]}_{\text{Return Relative to Benchmark}},$$

where $\bar{Q}$ represents the benchmark holding.

Assume the investor maximizes mean-variance utility over compensation

$$\max_Q \mathbb{E}\left[c(Q)\right] - \frac{\gamma}{2}\mathbb{V}\left[c(Q)\right].$$

One can show the optimal quantity demanded is

$$Q = \frac{\mathbb{E}\left[\tilde{D}\right] - P}{\gamma\mathbb{V}\left[\tilde{D}\right]\cdot(a+b)} + \frac{b}{a+b}\bar{Q},$$

and so the equilibrium price is

$$P = \mathbb{E}\left[\tilde{D}\right] - \left(\Theta - \frac{b}{a+b}\bar{Q}\right)\gamma\mathbb{V}\left[\tilde{D}\right](a+b).$$

The price multiplier is thus:

$$M = -\frac{\partial}{\partial\Theta}P = \gamma\mathbb{V}\left[\tilde{D}\right](a+b).$$

Note that the multiplier does not vary with shock size:

$$\frac{\partial}{\partial|\Theta|}M = 0.$$

Reformulating this setup in the notation of Section 2, we can write the cost function as

$$\begin{aligned}
C(Q,P) &= \mathbb{E}\left[Q\left(\tilde{D}-P\right)\right] - \mathbb{E}\left[c(Q)\right] + \frac{\gamma}{2}\left[\mathbb{V}\left[c(Q)\right] - \mathbb{V}\left[Q\left(\tilde{D}-P\right)\right]\right] \\
&= (1-a-b)\mathbb{E}\left[Q\left(\tilde{D}-P\right)\right] - b\mathbb{E}\left[\bar{Q}\left(\tilde{D}-P\right)\right] \\
&\quad + \frac{\gamma}{2}\mathbb{V}\left[\tilde{D}\right]\left[\left((a+b)^2-1\right)Q^2 - 2(a+b)bQ\bar{Q}\right]
\end{aligned}$$

Marginal cost is

$$\begin{aligned}
MC(Q,P) &\equiv \frac{\partial}{\partial Q}C(Q,P) \\
&= \left(\mathbb{E}\left[\tilde{D}\right]-P\right)(1-a-b) + \frac{\gamma}{2}\mathbb{V}\left[\tilde{D}\right]\left[2\left((a+b)^2-1\right)Q - 2(a+b)b\bar{Q}\right]
\end{aligned}$$

Thus, we have

$$\frac{\partial}{\partial\Theta}MC(\Theta,P) = \frac{\gamma}{2}\mathbb{V}\left[\tilde{D}\right]\left[2\left((a+b)^2-1\right)\right],$$

which is constant, so

$$\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta,P) = 0.$$

Similarly, we have

$$\frac{\partial}{\partial P}MC(\Theta,P) = -(1-a-b)$$

which is constant, so

$$\frac{\partial^2}{\partial|\Theta|\partial P}MC(\Theta,P) = 0.$$

A.4 Example of $\frac{\partial^2}{\partial|\Theta|\partial\Theta}MC(\Theta,P)<0$: Fixed Costs

Consider the following cost function:

$$C(Q,P)=\begin{cases}C & ,Q\neq 0\\ 0 & ,Q=0\end{cases}$$

which captures the idea of a fixed entry or adjustment (assuming the initial position is zero) cost.

One can show that the investor's quantity demanded is

$$Q=\begin{cases}\frac{\mathbb{E}\left[\tilde{D}\right]-P}{\gamma\mathbb{V}\left[\tilde{D}\right]} & ,\left(\mathbb{E}\left[\tilde{D}\right]-P\right)^2>2C\gamma\mathbb{V}\left[\tilde{D}\right]\\ 0 & ,\left(\mathbb{E}\left[\tilde{D}\right]-P\right)^2\leq 2C\gamma\mathbb{V}\left[\tilde{D}\right]\end{cases}$$

So the equilibrium price is

$$P=\begin{cases}\mathbb{E}\left[\tilde{D}\right]-\Theta\gamma\mathbb{V}\left[\tilde{D}\right] & ,|\Theta|>\sqrt{\frac{2C}{\gamma\mathbb{V}\left[\tilde{D}\right]}}\\ \left[-2C\gamma\mathbb{V}\left[\tilde{D}\right],2C\gamma\mathbb{V}\left[\tilde{D}\right]\right] & ,|\Theta|\leq\sqrt{\frac{2C}{\gamma\mathbb{V}\left[\tilde{D}\right]}}\end{cases}$$

where the price is indeterminate in the range $\left[-2C\gamma\mathbb{V}\left[\tilde{D}\right],2C\gamma\mathbb{V}\left[\tilde{D}\right]\right]$ when $0<|\Theta|\leq\sqrt{\frac{2C}{\gamma\mathbb{V}\left[\tilde{D}\right]}}$.

Thus, because there are discontinuities in price, the price multiplier $M=-\frac{\partial}{\partial\Theta}$ is effectively infinite for small $|\Theta|$ ($|\Theta|\leq\sqrt{\frac{2C}{\gamma\mathbb{V}\left[\tilde{D}\right]}}$), but then decreases to $M=\gamma\mathbb{V}\left[\tilde{D}\right]$ for large $|\Theta|$ ($|\Theta|>\sqrt{\frac{2C}{\gamma\mathbb{V}\left[\tilde{D}\right]}}$). Thus,

$$\frac{\partial}{\partial|\Theta|}M<0.$$

In the notation of Section 2, marginal cost is increasing in θ from ∞ to ∞ ($-\infty$ to $-\infty$) for $0<\Theta<\sqrt{\frac{2C}{\gamma\mathbb{V}\left[\tilde{D}\right]}}$ ($-\sqrt{\frac{2C}{\gamma\mathbb{V}\left[\tilde{D}\right]}}<0<\Theta$). But when Θ becomes large (small) enough and hits $\Theta=\sqrt{\frac{2C}{\gamma\mathbb{V}\left[\tilde{D}\right]}}$ ($\Theta=-\sqrt{\frac{2C}{\gamma\mathbb{V}\left[\tilde{D}\right]}}$), marginal cost switches from increasing to decreasing (i.e.

it decreases from infinty to 0 at $|\Theta| = \sqrt{\frac{2C}{\gamma \mathbb{V}[\tilde{D}]}}$). Thus,

$$\frac{\partial^2}{\partial|\Theta|\partial\Theta} MC(\Theta, P) < 0$$

at $|\Theta| = \sqrt{\frac{2C}{\gamma \mathbb{V}[\tilde{D}]}}$.

A.5 Example of $\frac{\partial^2}{\partial|\Theta|\partial P} MC(\Theta, P) > 0$: Costly Information Acquisition

A.5.1 Setup

There are two periods: $t = 1, 2$. Period $t = 1$ has two subperiods: $t = 1-$ and $t = 1+$. In $t = 1-$ information choice occurs. In $t = 1+$ portfolio choice occurs and the asset market clears. In $t = 2$ asset payoffs are realized.

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

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

The asset has exogenous, stochastic supply

$$\Theta \sim N(0, \sigma_\Theta^2),$$

which should be interpreted as the *residual* supply the representative investor faces: the total fixed, positive net supply minus the exogenous demand shocks of some noise traders. The exogenous risk-free rate is normalized to zero.

Agents and Preferences: There is a representative investor who chooses his portfolio at $t = 1+$ to maximize mean-variance utility over $t = 2$ (i.e. terminal) wealth

$$\max_Q \mathbb{E}_+ \left[Q\left(\tilde{D} - P\right)\right] - \frac{\gamma}{2}\mathbb{V}_+ \left[Q\left(\tilde{D} - P\right)\right] \tag{A.1}$$

Information: At $t = 1-$, the investor knows the supply Θ he will face at $t = 1+$, as well as the asset pricing parameters $\bar{D}, \sigma_\epsilon^2$, and σ_Θ^2.

However, he does not know η. He has the objectively correct prior and believes

$$\eta \sim N\left(0, \sigma_\eta^2\right). \tag{A.2}$$

The investor can pay a cost $C(G)$ to acquire signal a noisy signal s for η at $t = 1+$[1]:

$$\begin{aligned} s &= \eta + u, u \sim N\left(0, \sigma_u^2\right) \\ G &= \frac{\sigma_u^{-2}}{\sigma_u^{-2} + \sigma_\eta^{-2}}. \end{aligned} \tag{A.3}$$

G denotes the Bayesian gain of signal s. We assume the cost function $C(G)$ satisfies

$$C^{'}(G) > 0 \tag{A.4}$$

$$C^{''}(G) > 0 \tag{A.5}$$

Thus, the marginal cost of reducing uncertainty is positive ($C^{'}(G) > 0$) and increasing ($C^{''}(G) > 0$).

With signal s, the investor's posterior distribution is

$$\eta \sim N\left(G \cdot s, \frac{1}{\sigma_\eta^{-2}}(1-G)\right). \tag{A.6}$$

Since the investor does not know at $t = 1-$ what signal s he will observe at $t = 1+$, the investor chooses G to maximize *expected* utility at $t = 1-$ (which integrates over all possible realizations of s):

$$\max_G \mathbb{E}_- \left[\mathbb{E}_+ \left[Q\left(\tilde{D} - P\right)\right] - \frac{\gamma}{2}\mathbb{V}_+ \left[Q\left(\tilde{D} - P\right)\right]\right] - C(G) \tag{A.7}$$

The investor's beliefs are rational: his $t = 1-$ 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), information choices (G), and asset prices (P) such that:

1. The information choice G maximizes (A.7) given his prior beliefs about η and his knowledge of the supply Θ.

[1] As in e.g. Grossman and Stiglitz (1980); Kyle (1989); Van Nieuwerburgh and Veldkamp (2009, 2010); Kacperczyk et al. (2016); Han (2018)

2. The portfolio choice Q maximizes (A.1) given the investor's $t = 1+$ information set.

3. The asset market clears:
$$Q = \Theta. \tag{A.8}$$

A.5.2 Solving the Model

We solve the model backwards. We first fix the investor's $t = 1+$ information set $\mathcal{I}_+ = (s, G)$ and solve for the optimal quantity $Q(\mathcal{I}_+)$ that maximizes (A.1). We then impose market clearing (A.8) to solve for the equilibrium price. Finally, we solve for the investor's ex-ante optimal gain G at $t = 1-$ that maximizes (A.7) given the equilibrium $t = 1+$ quantity and price.

Portfolio Choice at $t = 1+$: Let $\bar{G}$ denote the equilibrium gain chosen in $t = 1-$. Then the $t = 1+$ information set is $\mathcal{I}_+ = \left(s, \bar{G}\right)$, where s is the random realization of (A.3). Given $\mathcal{I}_+$, the investor solves

$$\begin{aligned} &\max_Q \mathbb{E}_+\left[Q\left(\tilde{D} - P\right)\right] - \frac{\gamma}{2}\mathbb{V}_+\left[Q\left(\tilde{D} - P\right)\right] \\ \rightarrow Q\left(\mathcal{I}_+\right) = &\frac{\bar{D} + \bar{G}\cdot s - P}{\gamma\left(\sigma_\epsilon^2 + \frac{1}{\sigma_\eta^{-2}}\left(1 - \bar{G}\right)\right)}. \end{aligned} \tag{A.9}$$

Market Clearing at $t = 1+$: From market clearing (A.8) we have

$$P = \bar{D} + \bar{G}\cdot s - \Theta\gamma\left(\sigma_\epsilon^2 + \frac{1}{\sigma_\eta^{-2}}\left(1 - \bar{G}\right)\right). \tag{A.10}$$

That is, if supply equals its mean ($\Theta = 0$), then price equals expected terminal dividend $(\bar{D} + \bar{G}\cdot s)$. Exogenous supply shocks distort price away from this fundamental value.

Information Choice at $t = 1-$: At $t = 1-$ the investor knows what supply Θ he will face in $t = 1+$. He also knows the equilibrium distributions of quantity $Q\left(\mathcal{I}_+\right)$ and price P, which he takes as given.

Given the equilibrium quantity (A.9), the ex-ante information choice problem (A.7) becomes

$$\max_G \mathbb{E}_-\left[\frac{1}{2}\frac{\left(\mathbb{E}_+\left[\tilde{D}\right] - P\right)^2}{\gamma\mathbb{V}_+\left[\tilde{D}\right]}\right] - C(G).$$

Appendix A.5.4 shows the equilibrium first-order condition for the optimal G is:

$$\frac{\gamma}{2}\sigma_{\eta}^{2}\Theta^{2} = C^{'}(G). \tag{A.11}$$

A.5.3 Model Implications

More information acquisition for larger supply shocks. The following proposition, proven in Appendix A.5.5, establishes that the investor will choose a larger optimal gain when faced with supply shocks of larger magnitudes.

PROPOSITION 1 (More Information Acquisition for Larger Supply Shocks)**.** *Under the conditions* (A.4) *and* (A.5) *on $C(G)$, the optimal gain $G^{*}(\Theta)$ chosen to solve the investor's ex-ante information choice problem* (A.7) *is increasing in the size of the supply shock $|\Theta|$:*

$$\frac{\partial G^{*}(\Theta)}{\partial|\Theta|} > 0. \tag{A.12}$$

Intuitively, supply shocks that are larger in absolute magnitude distort price further away from fundamental value and thus yield larger expected returns in magnitude, which then raise the incentive to acquire information. More information acquisition at $t = 1-$ means the investor can more aggressively exploit the expected return in $t = 1+$ without exposing himself to a lot of subjective risk, which includes both volatility $\mathbb{V}_{+}[\epsilon]$ and uncertainty $\mathbb{V}_{+}[\eta]$.

Concave price impact Plugging in the equilibrium $G^{*}(\Theta)$ into the equilibrium price (

PROPOSITION 2 (Concave Price Impact)**.** *Under the conditions* (A.4) *and* (A.5)*, the per-unit price impact of the supply shock $M(\Theta)$ is decreasing in the magnitude of the shock $|\Theta|$:*

$$M(\Theta) = \gamma\left(\sigma_{\epsilon}^{2} + \frac{1}{\sigma_{\eta}^{-2}}\left(1 - G^{*}(\Theta)\right)\right) \tag{A.13}$$

$$\frac{\partial M(\Theta)}{\partial|\Theta|} < 0. \tag{A.14}$$

The intuition is that when faced with larger supply shocks that create larger expected returns, the investor will acquire more information, thereby reducing his uncertainty and making his demand more elastic. Since the investor more elastically absorbs the supply shock, the shock has less price impact.

Mapping to Section 2 Thus, in the notation of Section 2, we can write the cost function as

$$\begin{aligned} C(Q,P) &= \mathbb{V}_{+}\left[Q\left(\tilde{D}\right)\right] - \mathbb{V}\left[Q\left(\tilde{D}\right)\right] \\ &= Q^2\left(\sigma_\epsilon^2 + \sigma_\eta^2(1-G(P))\right) - Q^2\left(\sigma_\epsilon^2\right) \\ &= Q^2\sigma_\eta^2(1-G(P)) \end{aligned}$$

where $\mathbb{V}$ represents the variance taken under the objective measure, which only includes volatility ($\mathbb{V}[\epsilon]$) and not uncertainty about the expected dividend. Note that the equilibrium $G(P)$ depends on price P through the endogenous information choice (A.7).

So we have

$$\begin{aligned} MC(Q,P) &\equiv \frac{\partial}{\partial Q}C(Q,P) = 2Q\sigma_\eta^2(1-G(P)) \\ \frac{\partial}{\partial |Q|}MC(Q,P) &= \begin{cases} -2\sigma_\eta^2(1-G(P)) & ,Q<0 \\ 2\sigma_\eta^2(1-G(P)) & ,Q>0 \end{cases} \\ \frac{\partial^2}{\partial |Q|\partial P}MC(Q,P) = \frac{\partial^2}{\partial P\partial |Q|}MC(Q,P) &= \begin{cases} 2\sigma_\eta^2\frac{\partial}{\partial P}G(P) & ,Q<0 \\ -2\sigma_\eta^2\frac{\partial}{\partial P}G(P) & ,Q>0 \end{cases} \\ &> 0 \end{aligned}$$

The last line follows from Proposition 1 below, which establishes that G increases in $|\Theta|$ and so (since $Q=\Theta$ by market clearing)

$$\frac{\partial}{\partial P}G(P) = \begin{cases} >0 & ,Q<0 \\ <0 & ,Q>0 \end{cases}$$

A.5.4 Proof of (A.11)

Proof. At $t=1-$ the investor knows the supply Θ he will face in $t=1+$. He also knows the equilibrium distributions of quantity $Q\left(\mathcal{I}_{+}\right)$ and price P, which he takes as given.

Plugging in the equilibrium quantity (A.9) to the ex-ante information choice problem (A.7) yields

$$\max_G \mathbb{E}_{-}\left[\frac{1}{2}\frac{\left(\mathbb{E}_{+}\left[\tilde{D}\right]-P\right)^2}{\gamma\mathbb{V}_{+}\left[\tilde{D}\right]}\right].$$

Plugging in the posterior distribution for η at $t = 1+$ (A.6) and the equilibrium price (A.10) yields

$$\max_{G} \frac{1}{2\gamma} \frac{1}{\left[\sigma_\epsilon^2 + \frac{1}{\sigma_\eta^{-2}}(1-G)^2\right]^2} \left[\frac{\left(G-\bar{G}\right)^2}{\bar{G}\sigma_\eta^{-2}} + \Theta^2\gamma^2 \left[\sigma_\epsilon^2 + \frac{1}{\sigma_\eta^{-2}}\left(1-\bar{G}\right)^2\right]^2\right] - C\left(G\right).$$

Taking the first-order condition with respect to G and then imposing that in equilibrium $G = \bar{G}$ yields

$$\frac{\gamma}{2}\sigma_\eta^2\Theta^2 = C^{'}(G). \tag{A.15}$$

□

A.5.5 Proof of Proposition 1

Proof of Proposition 1. Recall the equilibrium first-order condition for the investor's ex-ante information choice problem is

$$\frac{\gamma}{2}\sigma_\eta^2\Theta^2 = C^{'}(G). \tag{A.16}$$

Since $C^{''}(G) > 0$ by condition (A.4), $C^{'}(G)$ is one-to-one and, hence, invertible. Thus, the optimal gain is

$$\begin{aligned} G^*(\Theta) &= (C^{'})^{-1}\left(\frac{\gamma}{2}\sigma_\eta^2\Theta^2\right) \\ &= (C^{'})^{-1}\left(\frac{\gamma}{2}\sigma_\eta^2\left|\Theta\right|^2\right) \end{aligned}$$

Since $C^{''}(G) > 0$ and $C(G)$ is twice continuously differentiable, $(C^{'})^{-1}$ is continuous and increasing (i.e. $\frac{\partial (C^{'})^{-1}(x)}{\partial x} > 0$. Thus,

$$\frac{\partial G^*(\Theta)}{\partial\left|\Theta\right|} = \left.\frac{\partial(C^{-1})^{-1}(x)}{\partial\left|\Theta\right|}\right|_{x=\frac{\gamma}{2}\sigma_\eta^2\left|\Theta\right|^2} \cdot \gamma\sigma_\eta^2\left|\Theta\right| > 0.$$

□

A.5.6 Proof of Proposition 2

Proof of Proposition 2. The multiplier is defined as

$$M\left(\Theta\right) = \gamma\left(\sigma_\epsilon^2 + \frac{1}{\sigma_\eta^{-2}}\left(1 - G^*\left(\Theta\right)\right)\right).$$

Taking the derivative with respect to the magnitude of the supply shock $|\Theta|$ yields:

$$\frac{\partial M(\Theta)}{\partial|\Theta|} = -\gamma\sigma_\eta^2 \frac{\partial G^*(\Theta)}{\partial|\Theta|} < 0,$$

since by Proposition 1

$$\frac{\partial G^*(\Theta)}{\partial|\Theta|} > 0.$$

□

B Supplemental Results for Demand Measurement

B.1 Description of Russell Banding Methodology Starting in 2007

Prior to 2007, 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 they're 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 they're 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.

B.2 Robustness in FIT Measurement

This section describes how we compute cleaned versions of flow-induced trading (FIT) to help alleviate measurement error concerns. Because our main result shows that price multipliers are smaller when the absolute value of FIT is larger, we need to alleviate the concern of overestimating large FIT values in the tails.

Where would mismeasurements come from? Under the simplest specification where we assume trades respond one-to-one to flows, the FIT of fund n in stock i is given by

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

From this perspective, there are two main mismeasurement concerns, and we tackle them in turn.

1. **Heterogeneous trading response to flows.** Trades respond less than one-to-one to flows, and this may be especially true when flows are large or when pre-existing positions are large. When a fund faces large inflows, it may use a larger fraction of the flows to buy new stocks, 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). Both of these considerations may lead us to over-estimate large FIT values.

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

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

   where $\text{Trade}_{n,i,t} = \frac{\text{SharesHeld}_{n,i,t}}{\text{SharesHeld}_{n,i,t-1}} - 1$ and $w_{n,i,t-1}$ is the lagged portfolio weight of stock i for fund n. To study heterogeneous responses, we sort the 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 $flow_{j,t} \in [-1\%, 1\%]$. Panel (a) of Figure B.1 plots the average flow by bins and show that they cover a large range from approximately -30% to +100%. Similarly, for each fund in each period, we sort its stock holdings by the existing portfolio weight 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 weighted regression with weights equal to $w_{n,i,t-1}$.[2]

Panel (b) of Figure B.1 plots the point estimates of $\beta_{b,f}$ for bins $b = 1, 10$, and 20. The results is consistent with the idea that, when facing large in flows, funds tend to trade less than one-to-one in stocks that they already have large holdings in. This effect is less pronunced for out flows, a finding that is consistent with Lou (2012).

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

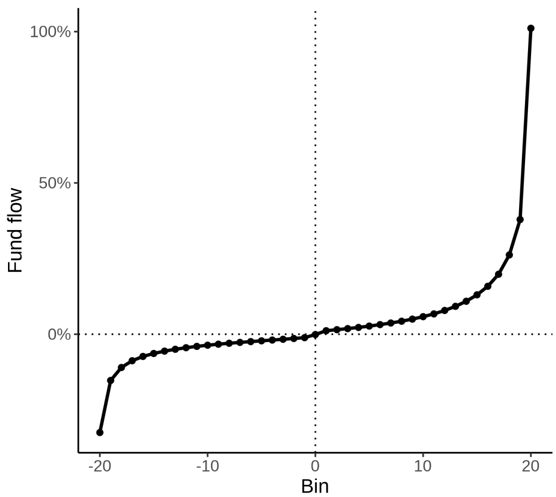
$$\text{FIT}_{i,t} = \sum_{\text{fund } n} \frac{\text{SharesHeld}_{n,i,t-1}}{\text{Share Outstanding}_{i,t-1}} \cdot \underbrace{\beta(w_{n,i,t-1}, f_{n,t})}_{\text{trade response}}$$

where the heterogeneous responses $\beta(w_{n,i,t-1}, f_{n,t})$ are based on the estimates in regression (B.1). Specifically, we first sort holdings into 20 bins by $w_{n,i,t-1}$ for each fund-quarter, and then apply a third-order polynomial-estimated curve based on the regression estimates.

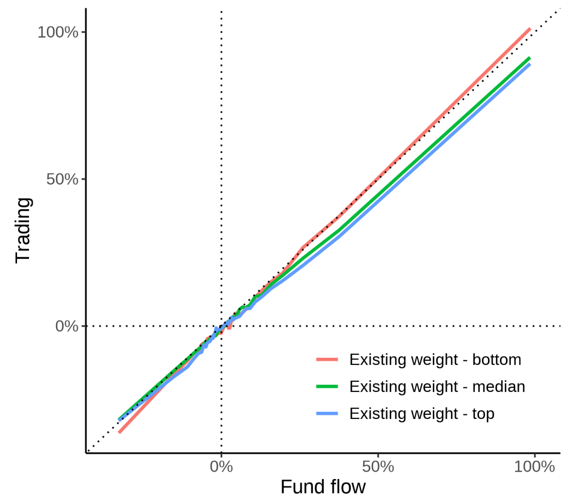
2. **Winsorize fund flows.** Another possible concern is that the extreme fund flows may be misestimated. To alleviate this concern, we recompute FIT after winsorizing 1%, 5%, or 10% of fund flows, with equal fraction of winsorization on each side. The winsorization thresholds are illustrated in Panel (c) of Figure B.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 meaures are not subject to mismeasured large flow values.

Panel (d) of Figure B.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 trade-to-flow response, and we plot it on the horizontal axis. On the vertical axis, the red line plots the average FIT after taking into account heterogeneous trade-to-flow 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.

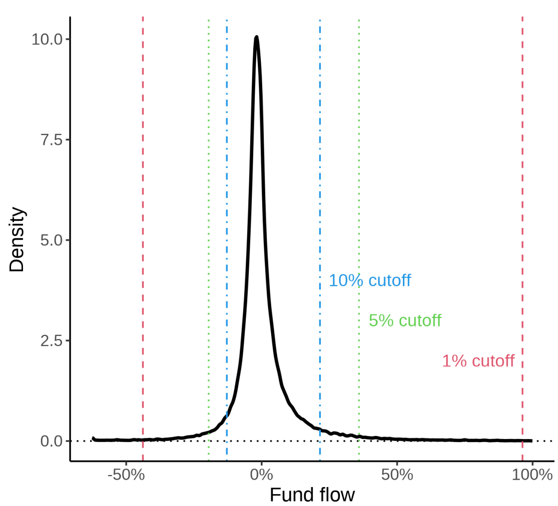
[2] To control for the fact that, even without flows, portfolio weights tend to mean-revert — that is, the largest (smallest) positions tend to be reduced (increased) subsequently — we estimate a first-stage regression where we regress the dependent variable on indicators of existing portfolio weight bin b. We use the resulting residuals as the dependent variable in regression (B.1).



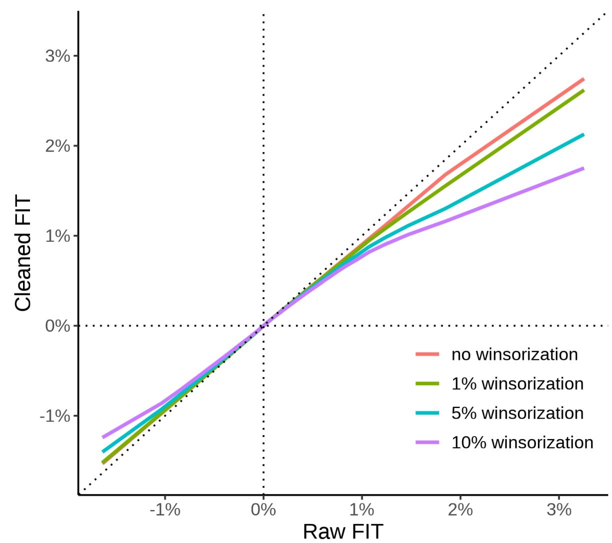
(a) Fund flow by bin



(b) Trade response to flow



(c) Fund flow winsorization



(d) Comparing FIT measures

Figure B.1. Cleaning the flow-induced trading (FIT) measure

Panel (a) plots fund 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, and the dashed diagonal line is the 45 degree line. Panel (c) plots the kernel denstiy 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.

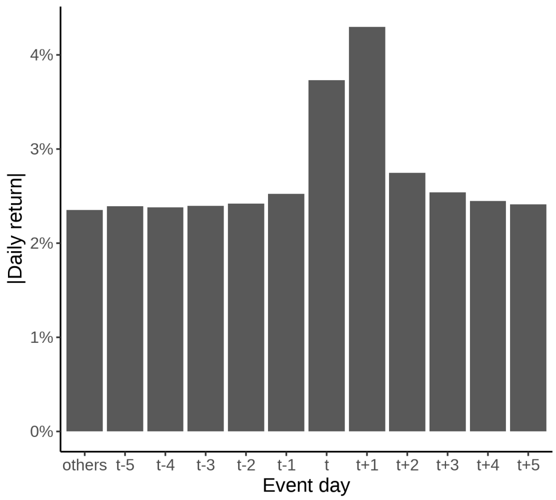
B.3 Cleaning the OFI Measure

This section describes how we refine the order flow imbalance (OFI) measure. We remove days with more information arrival based on three measures.

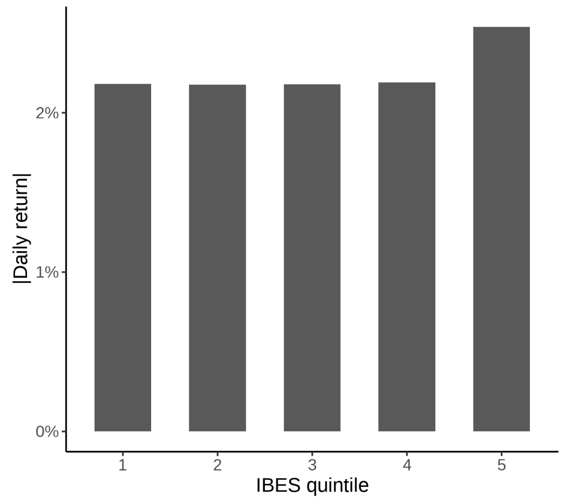
The first information measure is earnings announcement. We obtain quarterly earnings annoucement days from IBES and Compustat (variable "rdq"), and follow DellaVigna and Pollet (2009) to use the earlier of the two in case of disagreements. To gauge the amount and duration of information releases, Panel (a) of Figure B.2 plots the average absolute value of stock returns for days around earnings. The results are consistent with substantial information revelation on and also for up to two days around the earnings days.

The next two information measures are the number of analysts issuing updates in IBES and the number of news articles from Ravenpack. These two measures also have incremental power in explaining return variation. To see this, we sort days in each stock-quarter into quintiles based on these two measures, and plot the average absolute value of daily returns in Panels (b) and (c) of Figure B.2. The results suggest there is also information releases when these measures are in their top quintiles. It is worth explaining that both information measures are quite sparse, and especially for small firms, so much of the bottom four quintiles have zero updates based on both measures.

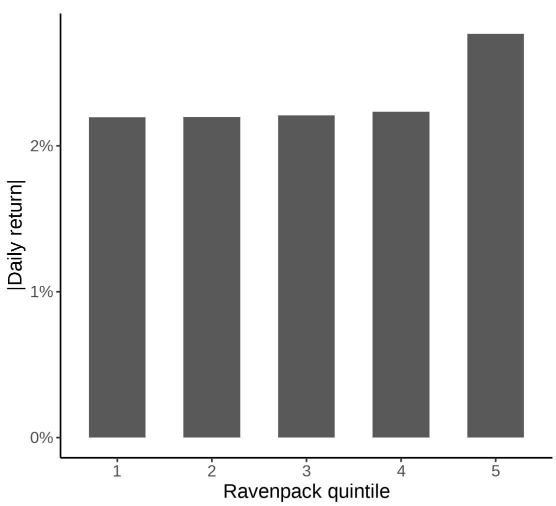
Based on the discussions thus far, we remove days with more information releases before computing quarterly OFI. Specifically, we remove all $t-3, ..., t+3$ days around each earnings day t, as well as days in the top quintiles based on IBES or Ravenpack updates. All results in the main paper use this refined OFI measure, but our results are robust to not doing this refinement. As an illustration, Panel (d) of Figure B.2 plots the interquartile range of the cleaned OFI measure against the raw OFI.



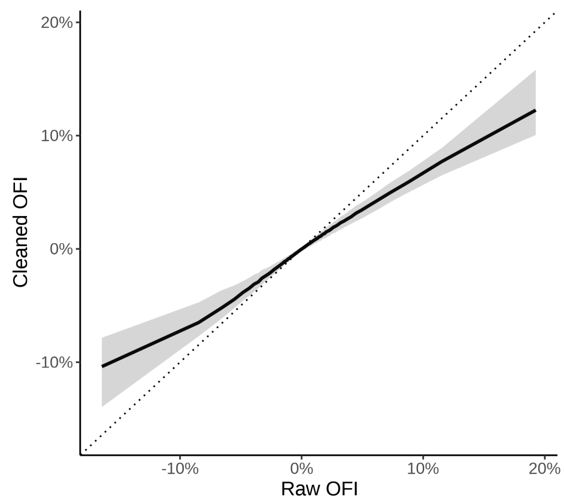
(a) Volatility around earnings



(b) Volatility by IBES



(c) Volatility by Ravenpack news



(d) Before vs. after cleaning

Figure B.2. Cleaning the order flow imbalance (OFI) measure

Panels (a) to (c) plot the average absolute value of daily returns by information indicators. Panel (a) examines the days around earnings events. In Panels (b) and (c), we sort days in each stock-quarter into quintiles based on the number of IBES analyst updates and the number of Ravenpack news articles, respectively. Panel (d) plots the interquartile range of the cleaned OFI where days with more information are removed against the raw OFI without cleaning.

C Additional empirical results

C.1 Supplements to Section 4

	Dependent variable: stock return $r_{i,t}$		
$d =$	BMI	FIT	OFI
	(1)	(2)	(3)
$d_{i,t}$	2.17***	4.57***	3.28***
	(0.68)	(0.61)	(0.14)
$d_{i,t} \times \lvert d_{i,t} \rvert$	−68.13***	−47.00*	−14.42***
	(25.37)	(25.70)	(0.68)
Controls	Y	Y	Y
Time FE	Y	Y	Y
Obs	9,910	544,662	529,619
Adj R^2	0.163	0.162	0.202
Marginal R^2 of demand	0.005	0.004	0.048

Table C.1. Interacted price impact regressions
We estimate panel regressions of stock returns on demand and demand interacted with its absolute value. The regressions control for commonly used stock characteristics and Fama-French 12 industry indicators as described in Section 3. We also control for time fixed effects, and cluster standard errors by time and stock. Column (1) reports results using the BMI measure of Pavlova and Sikorskaya (2023) and monthly returns. Columns (2) reports results using quarterly returns, whereas the demand is based on the FIT measure of Lou (2012) and the OFI measure in Li and Lin (2023), respectively. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

C.2 Price multipliers implied by event studies

To obtain estimates of price multipliers for large demand or supply shocks, we compute the multipliers implied by existing event studies. Specifically, companies announce buyback/issuance plans, and these events create contemporaneous price impacts at the announcement time. In typical event studies, researchers focus on the price movements on the day of the announcement, which helps isolate price movements induced by the announcements.

These events are associated with demand shocks that are much larger than existing studies for estimating price multipliers. For stock buybacks, we use estimates from Ikenberry et al. (1995) where the average demand shock is +6.64%, and the resulting average price effect is +3.54%, which implies a price multiplier of $3.54\%/6.64\% \approx 0.53$. For stock issuances, we use estimates from Asquith and Mullins Jr (1986) where the average demand shock is -8.71% of shares outstanding, but the average price impact is only -2.7%. The implied price multiplier is $2.7\%/8.71\% = 0.31$. Both of these are significantly smaller than that estimated from local demand shocks (e.g. Shleifer, 1986; Koijen and Yogo, 2019). While these studies are from an earlier period, subsequent studies have found approximately similar results (e.g. Eckbo et al., 2007). Overall, the event study-based results are also consistent with price multipliers being smaller for larger demand/supply shocks.

D Supplements to Nonlinear Asset Demand System

D.1 Derivation of Characteristics-Based Demand Function

Starting with investor i's optimization problem

$$\max_{\mathbf{Q}_i} \mathbf{Q}_i' \mathbb{E}_i \left[\mathbf{D}-\mathbf{P}\right] - \frac{\gamma_i}{2} \mathbf{Q}_i' \mathbb{V}_i \left[\mathbf{D}-\mathbf{P}\right] \mathbf{Q}_i - \pi_i \left\| \mathbf{Q}_i - \mathbf{B}_i \right\|_{\alpha_i}^{\alpha_i},$$

taking the firt-order condition with respect to $\mathbf{Q}_i$ yields

$$0 = \mathbb{E}_i \left[\mathbf{D}-\mathbf{P}\right] - \gamma_i \mathbb{V}_i \left[\mathbf{D}\right] \mathbf{Q}_i - \pi_i(\alpha_i - 1) \left| \mathbf{Q}_i - \mathbf{B}_i \right|^{\alpha_i - 1} sign(\mathbf{Q}_i - \mathbf{B}_i)$$

where $\left|\mathbf{Q}_i - \mathbf{B}_i\right|^{\alpha_i - 1} = [|Q_{i,n} - B_{i,n}|^{\alpha_i - 1}]_n$ and $\text{sign}(\mathbf{Q}_i - \mathbf{B}_i) = [\text{sign}(Q_{i,n} - B_{i,n})]_n$.

Plugging in the factor structure in cash flow expectations (16) yields:

$$\begin{aligned} 0 &= \boldsymbol{\mu}_i - \mathbf{P} - \gamma_i \left(\boldsymbol{\rho}_i \boldsymbol{\rho}_i' + \sigma_i^2 \mathbf{I}\right) \mathbf{Q}_i - \pi_i(\alpha_i - 1) \left|\mathbf{Q}_i - \mathbf{B}_i\right|^{\alpha_i - 1} sign(\mathbf{Q}_i - \mathbf{B}_i) \\ &= \boldsymbol{\mu}_i - \mathbf{P} - \gamma_i \boldsymbol{\rho}_i \underbrace{\boldsymbol{\rho}_i' \mathbf{Q}_i}_{\equiv c_i} - \gamma_i \sigma^2 \mathbf{Q}_i - \pi_i(\alpha_i - 1) \left|\mathbf{Q}_i - \mathbf{B}_i\right|^{\alpha_i - 1} sign(\mathbf{Q}_i - \mathbf{B}_i). \end{aligned}$$

Plugging in the expressions for $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$ yields for each stock n:

$$0 = \boldsymbol{\Phi}_i^{\mu'} \boldsymbol{x}_n + \phi_i^{\mu} - P_n - \gamma \left(\boldsymbol{\Phi}_i^{\rho'} \boldsymbol{x}_n + \phi_{i,n}^{\rho}\right) c_i - \gamma_i \sigma^2 Q_{i,n} - \pi_i(\alpha_i - 1) \left|Q_{i,n} - B_{i,n}\right|^{\alpha_i - 1}$$

Rearranging, for each stock n we have:

$$Q_{i,n} = \underbrace{\beta_{0,i}}_{\equiv -\frac{1}{\gamma\sigma^2}} P_n + \underbrace{\boldsymbol{\beta}_{1,i}'}_{\equiv \frac{\boldsymbol{\Phi}_{\mu,i}' - \gamma \boldsymbol{\rho}' \boldsymbol{Q}_i \boldsymbol{\Phi}_{\rho,i}'}{\gamma\sigma^2}} \mathbf{x}_n - \underbrace{\lambda_i}_{\equiv \frac{\pi\alpha}{\gamma\sigma^2}} \left|Q_{i,n} - B_{i,n}\right|^{(\alpha_i - 1)} \text{sign}(Q_{i,n} - B_{i,n}) + \underbrace{\epsilon_{i,n}}_{\equiv \frac{\phi_{\mu,i}' - \gamma \boldsymbol{\rho}' \boldsymbol{Q}_i \phi_{\rho,i}}{\gamma\sigma^2}}.$$

D.2 Calculating Counterfactual Prices

To calculate counterfactual prices, we adapt the algorithm from Koijen and Yogo (2019).

We want to solve the fixed point problem

$$\mathbf{me} = f(\mathbf{me}) \equiv \log\left(\sum_i A_i \mathbf{w}_i(\mathbf{me})\right) \tag{D.1}$$

where $\mathbf{me}$ is the vector of log market equity for each stock and $\mathbf{w}_i(\mathbf{me})$ is the vector of inside asset portfolio weights for investor i, which is a function of the log market equity of all stocks.

We solve this fixed point problem with the following algorithm with two nested loops

1. Given a guess for the log market equity vector at iterature k ($\mathbf{me}_k$), solve (20) numerically to find $\delta_{i,n,k}$ for each investor i and stock n and calculate the corresponding portfolio weight vector for each investor i $\mathbf{w}_i(\mathbf{me}_k)$.

2. Next, calculate the partial derivative of $\delta_{i,n,k}$ with respect to log market equity *at the current guess for the log market equity* $\mathbf{me}_k$:

$$\zeta_{i,n,k} \equiv \frac{\partial \delta_{i,n,k}}{\partial me_{n,k}} = \frac{\beta_{0,i,t} + (\alpha_i - 1)|K_{i,t} + \Delta \log \delta_{i,n,k} - \Delta me_{n,k}|^{\alpha_i - 1}}{1 + (\alpha_i - 1)|K_{i,t} + \Delta \log \delta_{i,n,k} - \Delta me_{n,k}|^{\alpha_i - 1}},$$

   where $\beta_{0,i,t}, K_{i,t}$, and $\alpha_{i,t}$ are fixed parameters from the estimated demand system in quarter t.

3. Now, holding $\zeta_{i,n,k}$ fixed, use Newton's method to solve (D.1) iteratively:

$$\begin{aligned}
\delta_{i,n,k,s} &= \delta_{i,n,k} + \zeta_{i,n,k}\left(\underbrace{me_{n,k,s}}_{\text{Inner loop guess}} - \underbrace{me_{n,k}}_{\text{Outer loop guess}}\right) \\
w_{i,n,k,s} &= \frac{\delta_{i,n,k,s}}{1 + \sum_m^N \delta_{i,m,k,s}} \\
\mathbf{me}_{k,s+1} &= \mathbf{me}_{k,s} + \left(\mathbf{I} - \frac{\partial f(\mathbf{me}_{k,s})}{\partial \mathbf{me}}\right)^{-1} (f(\mathbf{me}_{k,s}) - \mathbf{me}_{k,s})
\end{aligned} \tag{D.2}$$

   where we approximate the Jacobian with only its diagonal elements as in Koijen and Yogo (2019)

$$\frac{\partial f(\mathbf{me}_{k,s})}{\partial \mathbf{me}} = \operatorname{diag}\left(\min\left\{\frac{\sum_{i=1}^{l} \zeta_{i,n,k} A_i w_{i,n,k,s}(1 - w_{i,n,k,s})}{\sum_{i=1}^{l} A_i w_{i,n,k,s}}, 0\right\}\right).$$

   The minimum bounds the Jacobian away from $\mathbf{I}$ and prevents the step in the Newton's method (D.2) iteration from exploding. Iterate this step until convergence (usually fewer than 100 iterations are required). Note that k indexes the outer loop (steps 1 and 2), while s indexes this inner loop (step 3).

4. Return to step 1.

This nested-loop procedure reduces the number of times the $\delta_{i,n,t}$ must be solved for numerically, and so significantly reduces the computational cost. Usually fewer than 100 iterations of the outer loop are required.