Short versus Long-run Demand Elasticities in Asset Pricing*

Philippe van der Beck†

March 19, 2025

Abstract

The recent literature on demand-system asset pricing estimates investors' demand elasticities from static logit regressions, implying counterfactual experiments have constant impacts that do not revert over time. Using investors' trades at different horizons, I provide reduced-form evidence that elasticities increase significantly in the long run. To capture these dynamics structurally, I propose a partial adjustment model in which short and long-run elasticities differ due to portfolio inertia. I find that counterfactual price impacts are four times larger at quarterly horizons than in the long-run equilibrium. The model provides insights into return predictability via a novel stock-level measure of long-term reversal.

*I am grateful to Pierre Collin-Dufresne, Lorenzo Bretscher, Sam Hanson, Adi Sunderam, Mark Egan, Thomas Graeber, Semyon Malamud, Jean-Philippe Bouchaud, Paul Huebner, as well as the participants at the AFA 2023, EFA 2024, CFM Workshop on Market Impact, HEC Lausanne PhD Workshop, for helpful comments and valuable discussions. This paper supersedes a previous version titled "On the Estimation of Demand-based Asset Pricing Models", which was presented as part of the paper "Flow-Driven ESG Returns". I therefore also thank the seminar participants at Dauphine Finance PhD Workshop, Toulouse School of Economics, IESE Business School, Universitat Pompeu Fabra, Vrije Universiteit Amsterdam, Warwick Business School, University of Oxford Saïd Business School, Imperial College Business School, HEC Paris, Frankfurt School of Finance, Duke University, Carnegie Mellon University, Ohio State University, Boston University, Harvard Business School, University of Michigan, University of Toronto, Tuck School of Business, NYU Stern, University of Cologne, University of Hamburg, World Finance Conference Turin, and the SFI Research Days. The research was conducted under financial support from the Swiss National Science Foundation (SNSF).

†pvanderbeck@hbs.edu, Harvard Business School

Conflict of Interest Disclosure Statement

Philippe van der Beck, pvanderbeck@hbs.edu, Harvard Business School.
I have nothing to disclose.

1 Introduction

Detailed data on investor portfolio holdings has renewed interest in how investor heterogeneity, especially individual demand shocks, affects asset prices. If the aggregate demand curve for a financial security is downward sloping, investors' demand shocks can have a meaningful impact on equilibrium prices. The Demand System Approach to Asset Pricing developed by Koijen and Yogo (henceforth KY, 2019) allows incorporating holdings data into equilibrium asset pricing by estimating investor-specific demand curves. Investors' elasticity of demand with respect to the price (henceforth elasticity) lies at the heart of the demand system approach. It determines the equilibrium price impact of counterfactual experiments within the demand system. If the elasticity of demand is high, then equilibrium prices do not have to move a lot in order to induce investors to rebalance their portfolios, and the price impact of demand shocks is low.

While the idea that elasticities are horizon-dependent is pervasive throughout economics, the literature on asset demand systems has not yet confronted this issue. The standard approach in the demand-system asset pricing literature is to infer static elasticities from logit demand, thereby abstracting from potential horizon effects in the counterfactuals. A natural next step in this literature is to account for horizon effects and differentiate between short and long-run equilibrium counterfactuals. This paper addresses this gap, showing – both structurally and through a difference-in-differences estimator – that investors become significantly more elastic over longer horizons. Consistent across both estimations, I find that on average, investors' long-run elasticities are four times larger than their quarterly short-run elasticities. This implies that the impact of counterfactual experiments is four times larger at a quarterly horizon than in the long-term equilibrium. The patterns are in line with the slow-moving capital hypothesis (see Duffie (2010)) and imply that the price impact of counterfactual experiments slowly dissipates over time.

I set the stage by providing reduced-form evidence on increasing elasticities at longer horizons. I compute investors' trades, i.e. changes in their portfolio holdings, and estimate their elasticities over increasing trading horizons in a simple difference-in-difference setting. I find elasticities that increase monotonically at longer horizons. Investors trade 3 to 4 times more elastically over the long run than in the short run. The reduced-form estimates are closely linked to the standard logit regressions in the literature. The status quo is to estimate elasticities via cross-sectional regressions of portfolio holdings onto instrumented prices. The underlying intuition is that an investor is more price-elastic

in equilibrium if they overweight assets with lower prices. The cross-sectional identification from holdings does not take into account how investors *trade*, i.e., how they adjust their portfolios over time. The static model therefore does not speak to the horizon of investors' elasticities. The reduced-form estimates are first-difference versions of the static logit specification, with differences computed over varying trading horizons. The static logit elasticity identified from portfolio holdings in levels falls between the short and long-run elasticities identified from investors' trades. This suggests that static logit demand overstates how aggressively investors rebalance their portfolio in the short-run, and understates their rebalancing activity in the long-run.

Portfolio inertia is a natural candidate to explain differences in short and long-run elasticities. I first provide simple correlational evidence on slow-moving 13F portfolios. To this end, I construct passive buy-and-hold portfolios weights that use lagged shares at different horizons along with current prices. On average, an institution's passive portfolio from the previous quarter explains 89% of the variation in its current portfolio. The R^2 declines monotonically at longer horizons and reaches 63% with a 2-year lagged passive portfolio. The monotonically increasing rebalancing activity at longer horizons suggests that incorporating portfolio inertia in the demand estimation can deliver valuable insights regarding long-term elasticities and the price impact dynamics following demand shocks.

Motivated by the reduced-form evidence on increasing elasticities at longer horizons and portfolio inertia, I propose a simple dynamic extension to KY (2019) that allows differentiating between short and long-run effects in counterfactual experiments. I microfound a partial adjustment model following Gârleanu and Pedersen (2013) in which investors gradually shift towards a target portfolio due to trading costs. The optimal portfolio is a weighted average of the passive buy-and-hold portfolio and the target portfolio. The weight on the passive portfolio captures inertia, i.e., persistence in investor demand. I provide an instrumental variables estimator to address the endogeneity of past portfolios and current latent demand arising from persistent demand shocks.

In line with the reduced-form evidence, I find high inertia in institutional portfolios. The average investor's weight on the quarterly buy-and-hold portfolio is 0.84. This implies that investors partially trade towards their long-term target allocation with a quarterly adjustment speed of 0.16. In other words, every quarter investors close the gap between their current portfolio and their long-term target by 16%. Consistent with higher trading costs for bigger positions, large investors are more inert than small investors. The partial adjustment model implies that long-run elasticities are the interaction of short-run

elasticities and portfolio inertia. The short and long-run elasticities implied by the model are highly consistent with the reduced-form estimates. Slow portfolio adjustments lead to implied long-run elasticities that are around four times higher than in the short-run (quarterly). The AUM-weighted average short-run elasticity across all investors is 0.32, whereas the AUM-weighted average long-run elasticity is 1.2. There is also considerable heterogeneity in the long-run elasticity across investors. Over long horizons, large value investors are the most important liquidity providers with a long-run elasticity of 5.2. All results are robust to different instruments for lagged portfolios.

Elasticities that *increase* at longer horizons imply counterfactual price impacts that *decrease* over time. I illustrate this duality in a generalized setting that nests many workhorse asset pricing frameworks. In particular, I show that the equilibrium price effects of counterfactual experiments within a demand system can be approximated as the product of a demand shock and a multiplier matrix that is the inverse of the aggregate price elasticity of demand. If elasticities increase at longer horizons, the price multiplier decreases and the impact of counterfactuals dissipates over time. Correctly estimating elasticities and their dynamics therefore determines whether the demand-based approach produces realistic counterfactual price paths.

In the partial adjustment model, long-run elasticities – and consequently, long-run price impacts – are not directly estimated but inferred structurally from short-run elasticities and portfolio inertia. This addresses a key challenge in estimating long-horizon elasticities and price impact: the absence of plausibly exogenous instruments that remain relevant over long horizons. The difference in long-run and short-run price multipliers predicted by the model provides a novel stock-level measure of long-term reversal. The stock-specific reversal measure is given by the ownership-weighted increase in elasticities over long horizons. Intuitively, stocks in the habitat of long-term investors with high inertia exhibit strong reversal as these investors gradually rebalance their portfolios and become more elastic over time. I provide three concrete empirical applications of the dynamic elasticities implied by the partial adjustment model and the corresponding stock-specific reversal measure.

First, I show that model-based reversal strongly predicts which stocks exhibit greater price reversal following non-fundamental demand shocks. Coval and Stafford (2007) and Lou (2012) show that flow-induced price pressure by mutual funds significantly reverts over time. I sort stocks into portfolios based on the model-based reversal measure. For low reversal stocks, flow-induced price pressure decays slowly and remains significantly

positive for at least seven quarters. In contrast, for high reversal stocks, flow-induced price pressure fully reverses within three quarters.

Second, I simulate the predicted price path from counterfactuals implied by the partial adjustment model. To ensure a true out-of-sample test, I use demand shocks that have not been used to identify the model parameters. Out-of-sample, the dynamic model accurately predicts both the immediate effects of demand shocks and their long-term reversal, both qualitatively and quantitatively. In contrast, the static logit framework predicts constant price impacts, underestimating the short-run (quarterly) impact while overstating long-run price effects.

Last, I show that the model-based reversal measure contains strong incremental information beyond traditional cross-sectional return predictors. If momentum is driven partly by non-fundamental factors, stocks with higher model-based reversal should show weaker momentum because over time long-term elastic investors act contrarian against the initial demand shocks. When conditioning a standard momentum strategy on model-based reversal, momentum returns vanish for high reversal stocks, while low reversal stocks yield over 3% higher annualized momentum returns. I confirm these patters in Fama-MacBeth regressions. Conditioning on a reversal dummy doubles the coefficient on momentum for non-reversal stocks while eliminating it for reversal stocks. This underscores the incremental information embedded in investor portfolios and their rebalancing behavior more broadly.

Related Literature. The paper is related to three strands of literature. Most importantly, it directly relates to the recent and fast-growing literature on demand-based asset pricing initiated by KY (2019). This literature uses data on investors' portfolio holdings for different asset classes to estimate demand curves and, via the price elasticities of demand, link them to equilibrium prices (see Benetton and Compiani (2020), Koijen and Yogo (2020), Haddad et al. (2021), Han et al. (2021), Koijen et al. (2021), Fang et al. (2022), Coqueret (2022), Bretscher et al. (2024), Jiang et al. (2022), Darmouni et al. (2022), Tamoni et al. (2024), Jiang et al. (2024), Koijen et al. (2024), Jansen (2024), Jansen et al. (2024)). All of the papers mentioned above use the demand system approach with horizon-independent elasticities to conduct counterfactual experiments. Accounting for long versus short-run elasticities can provide valuable additional insights on how the price impact of counterfactual experiments is expected to unfold over time and to what extent it reverts. The partial adjustment model proposed in this paper

can be easily integrated into existing frameworks by controlling for lagged portfolios in the demand estimation. By effectively capturing static mandates that may be correlated with prices, this may furthermore sharpen the identification of existing studies. Huebner (2023) estimates a dynamic demand system and shows that if investors trade on momentum, short-term mispricing may worsen over time. The framework proposed in this paper allows separating long-run elasticities from momentum preferences and focuses on the former.

Second, the difference-in-difference estimator used to motivate the partial adjustment model is closely linked to the event-study evidence on price pressure from index reconstitutions (see e.g. Shleifer (1986), Wurgler and Zhuravskaya (2002), Chang et al. (2015), Pavlova and Sikorskaya (2022), Greenwood (2005)), mutual fund flows (see e.g. Edmans et al. (2012), Coval and Stafford (2007), Lou (2012), Wardlaw (2020), Schmickler (2020)) or dividend reinvestments (see e.g. Hartzmark and Solomon (2021) and Schmickler (2020)). All of the shocks introduced in these papers can essentially be used as instruments to estimate reduced-form elasticities via the difference-in-difference estimator. This lends additional flexibility to the demand-based approach as the elasticity estimates can be ex-ante verified in a host of different settings.

Finally, this paper contributes to the literature on slow-moving capital (e.g., Mitchell et al. (2007), Duffie (2010), Bogousslavsky et al. (2021)), gradual portfolio adjustment (e.g., Froot et al. (2001), Bohn and Tesar (1996), Bacchetta et al. (2023)) and investor inertia (e.g. Barber and Odean (2000), Calvet et al. (2009), Gabaix et al. (2024b)). I propose a partial adjustment model that can be estimated at the investor level, enabling the identification of both long-run and short-run elasticities from portfolio holdings data. Consistent with the theory on dynamic portfolio choice under transaction costs (e.g. Gârleanu and Pedersen (2013), Gârleanu and Pedersen (2016), Collin-Dufresne et al. (2020)), I find that larger investors with bigger positions trade more slowly than smaller investors.

The remainder of this paper is structured as follows. Section 2 provides a generalized framework that highlights the duality of demand elasticities and price impact in asset pricing. Section 3 describes the data. Section 4 estimates price elasticities in reduced form at different horizons. Section 5 provides reduced form evidence on portfolio inertia in 13F portfolios. Based on the reduced-form evidence, Section 6 microfounds and estimates the partial adjustment model. Section 7 introduces the new model-based measure of reversal. Section 8 provides applications and tests the model out-of-sample. Section 9 concludes.

2 The Role of Dynamic Demand Elasticities in Asset Pricing

Any asset pricing model is *demand-based* and implies an elasticity of demand with respect to the price that can in general be estimated from observable data on investors' portfolios. The increased availability of investor-level holdings data has revived the interest in estimating the slope of investors' demand curves and linking holdings to equilibrium prices. Neoclassical models generally imply high price elasticities of demand such that demand shocks have small equilibrium price effects.[1] Regardless of the precise magnitude of the elasticity, asset pricing models typically differ in their predictions about how this elasticity varies across horizons. Static models imply elasticities that are independent of the horizon.[2] Instead, dynamic models with slow-moving capital imply elasticities that increase over longer horizons.[3] These models offer tractable price dynamics following supply shocks, implying price impacts that revert over time as inert traders "wake up" or arbitrage/inventory risks decline. They typically solve for equilibrium prices directly and do not produce investor-specific demand curves in closed form. However, equilibrium price multipliers that *decline* over time are just the flip-side of demand elasticities that *increase* over time. Below I provide a generalized framework that highlights this duality.

A Portfolio Choice and Price Elasticities

There are N stocks indexed by $n = 1, ..., N$. Time is indexed by t. Shares outstanding of each stock are given by Q^*, normalized to 1 so that the endogenous total market equity of each stock is given by its price $P_t(n)$. There are I investors indexed by $i = 1, ..., I$ who choose an optimal number of shares $Q_{i,t}(n)$ in each stock. For now, I am assuming that the investors' optimal portfolio $Q_{i,t} \in \mathbb{R}^N$ is some fund-specific function $f_i(\cdot)$ of the vector of stock prices P_t and a collection of other exogenous observable and unobservable demand shifters X_t (such as e.g. expected cash flows, the assets under management,

[1] See the calibrations in Petajisto (2009), Gabaix and Koijen (2023), and Gabaix et al. (2024b). The negligible role of demand shocks in standard models is at odds with extensive empirical evidence from index inclusions mutual fund flows, index reconstitutions, benchmarking (Pavlova and Sikorskaya (2022)), and dividend reinvestments, suggesting that investors have downward-sloping (or inelastic) demand curves. See Wurgler and Zhuravskaya (2002), Lou (2012), Chang et al. (2015), and Hartzmark and Solomon (2021) for respective examples.

[2] See e.g. Wurgler and Zhuravskaya (2002), Petajisto (2009), Pavlova and Sikorskaya (2022), Kashyap et al. (2023), Koijen et al. (2024).

[3] See e.g. Grossman and Miller (1988), Greenwood (2005), Duffie (2010), Bogousslavsky (2016)).

interest rate, risk or investment constraints)

$$Q_{i,t} = f_i\big(P_t, X_t, Q_{i,t-1}\big) \tag{1}$$

This general specification allows plugging in a demand curve which matches observable data on portfolio choice. For example, in a static CARA setting $f_i(\cdot)$ is linear in prices and $\partial f_i/\partial Q'_{i,t-1} = 0$. Via the lagged holdings term $Q_{i,t-1}$ the optimal portfolio possibly depends on lagged characteristics and prices. This introduces path dependence and can be motivated by portfolio adjustment costs. For example, in Gârleanu and Pedersen (2013) $f_i(\cdot)$satisfies a partial adjustment model given by $Q_{i,t} = \rho_i Q_{i,t-1} + g_i(P_t, X_t)$. KY (2019) use an exponential function in a static environment with $\partial f_i/\partial Q'_{i,t-1} = 0$.[4] Balasubramaniam et al. (2021) and Betermier et al. (2022) opt for a linear specification. Importantly, $f_i(\cdot)$ pins down an investor's $N \times N$ instantaneous (short-run) elasticity matrix

$$\zeta_{i,t} = -\frac{\partial \log Q_{i,t}}{\partial \log P'_t} \tag{2}$$

which has own-price elasticities $-\frac{\partial \log Q_{i,t}(n)}{\partial P_t(n)}$ on the diagonal and cross-price elasticities $-\frac{\partial \log Q_{i,t}(n)}{\partial P_t(m)}$ on the off-diagonal elements. Appendix A.1 shows that for a CARA utility investor with risk aversion γ_i and a covariance matrix of future cashflows Σ the elasticity matrix is proportional to $(\gamma_i \Sigma)^{-1}$. Risk-averse arbitrageurs require larger price concessions to move away from stocks which contribute more to the risk of the arbitrage portfolio. If investors are inert $\partial f_i/\partial Q'_{i,t-1} \neq 0$, their elasticity is horizon-dependent. The long-run elasticity is the cumulative response to price changes over h periods and is given by

$$\zeta_{i,t}^{(h)} = \zeta_{i,t}\left(\mathbf{I} + \sum_{s=1}^{h}\left(\frac{\partial f_i}{\partial Q'_{i,t-1}}\right)^s\right) \tag{3}$$

See Appendix A.2 for a proof. When $\frac{\partial f_i}{\partial Q'_{i,t-1}} = \rho_i \mathbf{I}$, the long-horizon elasticity simplifies to $\zeta_{i,t}^{LR} = \frac{1}{1-\rho_i}\zeta_{i,t}$ as $h \to \infty$. For example, in Gârleanu and Pedersen (2013), transaction costs cause $\rho_i > 0$, which leads to portfolio inertia and long-run elasticities that exceed the instantaneous elasticity $\zeta_{i,t}$ by a factor $\frac{1}{1-\rho_i}$.

[4] They provide a suggestive micro-foundation of log-linear portfolio weights by using the fact that $w(n) = 1 + X(n) + \frac{1}{2}X(n)^2 ... \approx e^{X(n)}$.

B Market Clearing and Price Multipliers

Market clearing implies that all funds must jointly hold the market portfolio $\sum_{i=1}^{I} f_i(\cdot) = Q^*$. Equilibrium prices are only available in closed form for very specific demand curves. For example, linear CARA-normal demand implies the CAPM (see Appendix A.1). For general demand curves – like the exponential demand specification in KY (2019) – the price impact of counterfactual experiments can be first-order approximated as the product of a demand shock and the market's aggregate price elasticity of demand. To this end, assume that there is an exogenous shock ΔX_t that leads to a change in aggregate investor demand of $\Delta d_t \in \mathbb{R}^N$ (expressed in percent of shares outstanding).

PROPOSITION 1 *A first-order approximation of the equilibrium return $\Delta p_{t:t+h}$ between t and $t+h$ due to the demand shock Δd_t is given by*

$$\Delta p_{t:t+h} = \mathcal{M}_t^{(h)} \Delta d_t \tag{4}$$

where $\mathcal{M}_t^{(h)} \in \mathbb{R}^{N \times N}$ is a price impact matrix given by the inverse of the ownership-weighted average elasticity over horizon h

$$\mathcal{M}_t^{(h)} = \left(\sum_{i=1}^{I} diag(Q_{i,t}) \zeta_{i,t}^{(h)} \right)^{-1}. \tag{5}$$

See Appendix A.3 for a proof. Proposition 1 states that the impact of demand shocks on the cross-section of realized returns is given by the market's aggregate elasticity of demand. $\mathcal{M}$ is the cross-sectional pendant to the scalar multiplier in Gabaix and Koijen (2023) and will henceforth be referred to as the multiplier matrix.[5] Elasticities that increase over longer horizons h imply that the price impact of demand shocks reverts over time as the multiplier decreases. Ultimately, elasticities and therefore the multiplier matrix are pinned down by the specific choice of $f_i(\cdot)$. In a static CARA-normal setting the multiplier matrix is proportional to $\gamma_i \Sigma$. Greenwood (2005) shows that in a multi-period CARA setting the price impact of demand shocks decreases over time as cash-flow risk declines closer to the liquidation time of the securities. Appendix A.4 shows that any counterfactual experiment in KY (2019) can in general be restated in the form of an aggregate demand shock Δd_t expressed in percent of shares outstanding. A first-order

[5]The more price-elastic investors are (i.e., the larger the diagonal elements in ζ_i), the less prices have to move to accommodate the exogenous demand shock. Cross-price elasticities drive the off-diagonal elements in $\mathcal{M}$ and are responsible for spill-over effects to other stocks.

approximation of the price impact of counterfactual experiments in logit demand systems is given by equation (4) with a static (instantaneous) elasticity $\zeta_{i,t}$.[6] Effectively, the static framework compares two equilibria before and after the shock and does not speak to the dynamics of price path following counterfactual experiments. Depending on the way the elasticities and therefore the multipliers are estimated, the price impact captures long versus short-run dynamics.

3 Data and Variable Construction

I obtain institution-level share holdings $S_{i,t}(n)$ from 1990 to 2024 from Thomson's Institutional Holdings Database (s34 file). The data on quarterly stock holdings are subsequently merged with price and fundamentals data from CRSP and Compustat.[7] Institution-level and mutual fund portfolio weights $w_{i,t}(n)$ are constructed as the dollar holdings in each stock (price $P_t(n)$ times shares held $Q_{i,t}(n)$) divided by their assets under management $A_{i,t}$. An institution's assets under management are given by the sum of its dollar holdings. The subscripts t indicate the report date of the 13F filing. Monthly data on mutual funds' net returns and total net assets, as well as other fund-specific characteristics, are obtained from the CRSP survivorship-bias-free mutual fund database. Fund flows relative to lagged assets are denoted as $f_{i,t} = \frac{F_{i,t}}{A_{i,t-1}}$. In line with the theory, I normalize by shares outstanding such that a stock's price and its market capitalization coincide. Thus the number of shares held is equivalent to the fraction of ownership $Q_{i,t}(n) = \frac{S_{i,t}(n)}{\text{Shrout}_t(n)}$, and stock splits do not contaminate changes in shares held from one quarter to the next. I define $\delta_{i,t}(n) \equiv \log \frac{w_{i,t}(n)}{w_{i,t}(0)}$ where $w_{i,t}(0)$ is the weight in the outside asset as in KY (2019). I denote log quantities and market equities using lowercase, i.e. $p_t(n) = \log P_t(n)$ and $q_{i,t}(n) = \log Q_{i,t}(n)$. Δ denotes quarterly changes. Changes beyond single quarters are denoted by Δ^h, where h is the horizon (in quarterly units) over which the change is computed. Percentage changes in prices and quantities are hence given by $\Delta^h p_t(n) = \log P_t(n) - \log P_{t-h}(n)$, $\Delta^h q_{i,t}(n) = \log Q_{i,t}(n) - \log Q_{i,t-h}(n)$ and $\Delta^h \delta_{i,t}(n) = \delta_{i,t}(n) - \delta_{i,t-h}(n)$.

[6] For example, Koijen et al. (2024) assess the impact of a shift towards passive management by redistributing assets from active to passive funds. Han et al. (2021) evaluate the price impact of mutual fund risk shifting by redistributing underperforming funds' assets. Bretscher et al. (2024) conduct a battery of counterfactuals for the corporate bond market by changing the wealth distribution across investors A_i and bond characteristics X. Jiang et al. (2024) decompose the US net foreign asset position by iteratively changing its underlying determinants and computing the counterfactual equilibrium. The counterfactual price change is – at least to a first order – driven by $\mathcal{M}$.

[7] See KY (2019) for details on the construction of the database.

4 Reduced-Form Evidence on Horizon-Specific Elasticities

Before diving into the elasticity estimation it is worth noting that virtually all of the reduced-form evidence for inelastic demand is cast in the form of regressions of price changes onto some exogenous demand shift by a specific investor group, such as index demand by passive funds or flow-induced demand by mutual funds. By changing the horizon of the price change, the resulting coefficient captures short versus long-run price multipliers. In this spirit, I propose a reduced-form estimator for investor-specific elasticities (the inverse of multipliers) using investors' trades at different horizons.

A A simple difference-in-difference estimator

The most general reduced-form estimator, that directly emerges from the definition of the elasticity matrix in equation (2), is evaluating the percentage trade $\Delta^h q_i(n)$ as a response to a percentage change in the price $\Delta^h p(n)$ of all other stocks over different horizons h. I am omitting time t subscripts to avoid notational clutter. Appendix B.1 shows that under logit demand, own-price elasticities are constant across stocks and cross-price elasticities are subsumed by an investor-specific constant c_i. Unless stated otherwise, $\zeta_i^{(h)}$ henceforth denotes an investor's scalar own-price elasticity. These elasticities can be estimated in a simple cross-sectional regression for each investor i.[8]

$$\Delta^h q_i(n) = -\zeta_i^{(h)} \Delta^h p(n) + c_i + d_i(n) \tag{6}$$

There are good reasons as to why cross- and own-price elasticities may differ across stocks. For example, a simple mean-variance portfolio implies heterogeneous cross-elasticities driven by covariances across stocks. Note, that the focus of this paper does not lie on capturing the specific substitution patterns across different stocks, but on assessing how investors' average elasticity changes at longer time horizons. Chaudhary et al. (2023) provide a detailed analysis of heterogenous substitution patterns in corporate bonds. I leave the analysis of horizon effects in cross-elasticities up to future research.

[8]Technically, $\zeta_i^{(h)}$ measures the difference between own- and cross-price elasticity. However, for diversified portfolios, this difference is negligible as individual portfolio weights are small.

B Identification

Identifying how investors change their holdings, i.e., trade, as a response to changes in the prices, requires instruments for price changes, i.e., returns. To construct exogenous return variation at different frequencies, I follow Lou (2012) and use flow-induced trades by mutual funds.[9] To the extent that fund flows are correlated with the funds' respective portfolio tilts, flow-induced trades may contain fundamental information and violate the exclusion restriction. I, therefore, construct flow shocks $f_{i,t}^{\perp}$ by orthogonalizing flows with respect to fund-level characteristics. Investor-specific elasticities can then be obtained in a simple two-stage least squares procedure. Because flow-induced trading is a weak instrument over the small subsets of stocks held by most investors, I estimate the first stage jointly across all stocks in a two-sample instrumental variables estimation.[10] Appendix B.2 and B.3 provide further details on the instrument construction and estimation.

C Increasing Elasticities at Longer Horizons

Table 1 reports the estimated investor-specific elasticities for horizons h ranging from 1 quarter up to 4 years. The first column reports the AUM-weighted average elasticity across all investors. At a quarterly horizon, the average elasticity is 0.58 and increases monotonically to 1.68 at the 4-year horizon. Similarly, the median elasticity across investors increases monotonically from 0.47 at a quarterly horizon to 2.5 at long horizons. At the 90th percentile among investors, long-run elasticities are over 6 times larger than their short-term counterparts. Unsurprisingly, the AUM-weighted average elasticity approximately equals the inverse of the first-stage coefficient. Consistent with the notion of increasing elasticities, the first-stage multiplier decreases monotonically over longer horizons. Generally, finding strong instruments to estimate *long-run* elasticities is challenging: If a demand shock significantly affects returns over very long horizons, it probably contains fundamental information and therefore violates the exclusion restriction. However, the primary goal of this section is not to pin down investors' exact elasticities, but to provide simple reduced-form evidence of increasing elasticities at larger horizons. The structural model introduced in Section 6 addresses the challenge of weak long-run instruments by identifying horizon-dependent elasticities from portfolio inertia.

[9] The reduced-form literature on price multipliers has proposed several alternative exogenous demand shifts that could be used to instrument for price changes, such as index reconstitutions (see e.g. Chang et al. (2015), Pavlova and Sikorskaya (2022)) and, more recently, signed order flow aggregated to lower frequencies (Li and Lin (2022)).

[10] See e.g. Angrist and Krueger (1992), Huebner (2023), and Koijen et al. (2024).

Overall, the reduced-form results paint a consistent picture of elasticities increasing substantially over longer horizons, with investors exhibiting roughly 3 to 4 times greater elasticities in the long run compared to the quarterly frequency.

[Table 1 about here.]

D Relation to Structural Estimates from Logit Demand

In a conceptual deviation from the event-study evidence for inelastic demand, the literature on demand systems in asset pricing identifies elasticities structurally from portfolio holdings in *levels*. Appendix B.4 shows that the standard logit specification can be expressed as a simple cross-sectional regression given by

$$q_i(n) = -\zeta_i p(n) + c_i + d_i(n). \tag{7}$$

The reduced-form estimator (6) for horizon h is a straightforward first-difference version of the level specification, with differences computed over h periods. Flow-induced trades impact returns but are not significantly related to market equities in levels, even over long horizons. I, therefore, instrument endogenous market equity $p(n)$ with KY's mandate-based instrument. The bottom row of Table 1 reports the elasticities obtained from the level specification with an AUM-weighted average elasticity of 0.73. Despite using different estimators and instruments, the logit and reduced-form elasticities are both low and comparable in magnitude. This is another piece of evidence that institutional demand is indeed far more inelastic than implied by neoclassical frameworks (see Gabaix and Koijen (2023)). Nevertheless, the level regressions are static and capture an average horizon-independent effect that lies between the short and long-run elasticity from the reduced-form estimation.[11]

5 Reduced-Form Evidence on Portfolio Inertia

Differences in short and long-run elasticities can come from various sources. The purpose of this paper is not to precisely attribute increasing elasticities to a single channel. However, a compelling explanation is inertia in the portfolios of large institutional investors.

[11] Note, however, that the estimates are not directly comparable as they rely on different instruments. The partial adjustment model proposed below estimates both horizon-specific and static elasticities with the same instrument and confirms that the static logit estimate indeed lies between the short and long-run elasticity.

Portfolio inertia refers to the notion that investors' portfolios tend to be "sticky". Investors stick to past allocation decisions (or default options) over long horizons and infrequently adjust their portfolios. Note, that this refers to the capital allocation in terms of shares, not portfolio weights. Because market prices vary, portfolio weights may strongly change over time without any active trading decisions in terms of explicit share purchases and sales.

I provide simple reduced-form correlations that highlight just how inert 13F institutions are in the short run. To this end, I define an investor's passive portfolio $w^p_{i,t}(n)$ as the buy-and-hold portfolio the investor would hold in the absence of active trading between $t-1$ and t. The passive portfolio is constructed using current prices but lagged share holdings, $w^p_{i,t}(n) = \frac{Q_{i,t-h}(n)P_t(n)}{\sum_n Q_{i,t-h}(n)P_t(n)}$. I construct the passive portfolio using lagged shares for h ranging from 1 to 8 quarterly lags. For each institution-quarter pair with over 100 holdings, I run the simple cross-sectional regression

$$w(n) = \alpha + \beta w^p(n) + \epsilon(n). \tag{8}$$

The R^2 from this regression are strikingly high and monotonically decrease at longer lags for the passive portfolio. Figure 1 plots the distribution of R^2 for all investor-quarter pairs. On average, the passive portfolio of the previous quarter explains 89% of an institutions current portfolio. At an annual lag, the passive portfolio still explains 74% of the variation. The explained variation drops to 63% at the 2-year lag. While these figures focus on portfolio rebalancing behaviour at the intensive margin, the high explanatory power of lagged holdings for current portfolios is robust to accounting for portfolio inclusions and exclusions, i.e., extensive trades.[12] The fact that current portfolios are well explained by passive portfolios constructed from lagged shares implies that investors do not strongly rebalance their portfolios at short horizons. Their rebalancing activity monotonically increases at longer horizons, which suggests that incorporating portfolio inertia in the demand estimation can deliver valuable insights regarding the long-term price dynamics of demand shocks.[13]

[Figure 1 about here.]

[12] See Figure D.4 in the Appendix.

[13] The simple univariate regressions cannot establish causal evidence of investor inertia. The high explanatory power of lagged portfolios may be driven by persistent trading signals. Causal identification of portfolio inertia requires exogenous variation in lagged portfolios unrelated to unobservable slow-moving trading signals. The simple univariate regressions furthermore do not allow separating inertia from inelasticity. The partial adjustment model introduced below addresses these concerns.

6 Inertia and Dynamic Elasticities: A Structural Approach

Motivated by the horizon-dependent elasticities and the high correlations of lagged and current portfolios, I propose a simple dynamic extension to the logit specification in KY (2019). This allows incorporating horizon effects in the counterfactual experiments of existing and future applications of the demand-system approach. Beyond improved counterfactuals, the structural approach has three other key benefits. First, long-run elasticities are not directly estimated but inferred from short-run elasticities combined with portfolio inertia. This mitigates a key challenge in estimating long-horizon elasticities: The lack of plausibly exogenous instruments that are relevant at long horizons. Second, the first-difference estimator can obscure valuable insights, particularly for slow-moving holdings. Intuitively, many long-run investors respond slowly to new investment opportunities. For example, a value investor may rebalance annually or gradually trade towards their aim portfolio. Over-differencing their slow-moving portfolio holdings may remove valuable long-run information contained in the data causing the reduced-form estimator to fit noise from quarters in which the institution only partially rebalances or does not trade at all. Last, the reduced-form estimator can only be applied to investors with a large time series of available trades. Instead, the structural estimator can be estimated with little pooling across time and investors and, therefore, uncovers important heterogeneity in price impact reversal across stocks and time.

A A Partial Adjustment Model based on Transaction Costs

The notation is the same as in the generalized model in Section 2. Investor i chooses an optimal portfolio $w_{i,t} \in \mathbb{R}^N$ at time t subject to her subjective beliefs about expected returns $\mu_{i,t}$ and covariances $\Sigma_{i,t}$. Let $w^p_{i,t} \in \mathbb{R}^N$ denote the passive buy-and-hold portfolio the investor would hold in the absence of active portfolio rebalancing between $t-1$ and t. The corresponding log weight relative to the outside asset is denoted by $\delta^p_{i,t}(n) \equiv \log \frac{w^p_{i,t}(n)}{w_{i,t}(0)}$. The investor's *active* transactions between $t-1$ and t are given by $w_{i,t} - w^p_{i,t}$. Appendix A.5 shows that if the investor faces quadratic transaction that are proportional to $\Sigma_{i,t}$, the dynamic demand curve of the investor is a partial adjustment model: The optimal portfolio is a weighted mix of the desired long-run target portfolio $\delta^*_{i,t}$ and the passive buy-and-hold portfolio $\delta^p_{i,t}$,

$$\delta_{i,t} = \rho_{i,t}\delta^p_{i,t} + (1-\rho_{i,t})\delta^*_{i,t} \tag{9}$$

where the weight on the passive buy and hold portfolio $\rho_{i,t} \in (0, 1]$ captures portfolio inertia, which is driven by transaction costs and risk aversion. In asset pricing, trading costs are a natural micro-foundation for slow portfolio adjustments (Gârleanu and Pedersen (2013)). However, they are just *one* possible reason for portfolio inertia, which could also originate from both behavioral factors, such as cognitive constraints (Enke and Graeber (2023)). One could therefore conceive alternative microfoundations for (9) based on, for example, inattention. Portfolio inertia prevents the investor from fully adjusting to her target portfolio $\delta^*_{i,t}(n) \equiv \log \frac{w^*_{i,t}(n)}{w_{i,t}(0)}$ where

$$w^*_{i,t} = \frac{1}{\gamma_i} \Sigma^{-1}_{i,t} \mu_{i,t}. \tag{10}$$

When expected returns and covariance factor loadings are linear in characteristics, the target portfolio can be expressed as a log-linear function of log market equity $p_t(n)$ and a demand shifter $d_{i,t}(n)$. Formally,

$$\delta^*_{i,t}(n) = \beta_{i,t} p_t(n) + d_{i,t}(n) \tag{11}$$

The log-linearity of the partial adjustment model stems from a relatively strong coefficient constraint on the factor loadings driving returns and covariances (see Appendix A.5), which is primarily made for empirical convenience. Portfolio weights tend to be log-normally distributed in the cross-section. Furthermore, the log-linearity allows for a convenient separation of elasticities and inertia. Calvet et al. (2009) propose a similar specification for the cross-section of household rebalancing behavior. In an earlier version of their paper, Gabaix et al. (2024a) introduce a partial adjustment model for capturing inertia in the time series of household demand. Similar partial adjustment models have been employed in capital structure regressions by Fama and French (2002) and Kayhan and Titman (2007).

Differentiating the target portfolio $\delta^*_{i,t}(n)$ with respect to log prices one can solve for the long-run equilibrium elasticity

$$\zeta^{LR}_{i,t} \approx 1 - \beta_{i,t}, \tag{12}$$

which captures the extent to which the investor bets against deviations from fundamental value in the long run. However, the investor only partially trades towards this long-run elasticity. Differentiating the actual portfolio $\delta_{i,t}(n)$ with respect to log prices yields the

short-run elasticity

$$\zeta_{i,t}^{SR} \approx (1 - \beta_{i,t})(1 - \rho_{i,t}). \tag{13}$$

Therefore the long-run elasticity exceeds the short-run by the partial adjustment factor $(1 - \rho_{i,t})^{-1}$. This approximation holds for diversified portfolios with small $w_{i,t}(n)$. Appendix A.6 provides derivations and exact expressions including the full elasticity matrix with the cross-elasticities implied by the budget constraint.

B Estimating the Partial Adjustment Model

The passive portfolio can be decomposed into lagged ownership relative to shares outstanding, $q_{i,t-1}(n)$, and current market equity. Formally, $\delta_{i,t}^{p}(n) = q_{i,t-1}(n) + p_t(n) - h_{i,t}$ where $h_{i,t} = \log(A_{i,t} w_{i,t}(0))$ are the log dollar holdings in the outside asset. For simplicity, I assume that the holdings in the outside asset are constant between $t-1$ and t. Any portfolio-level growth is subsumed by time-fixed effects in the empirical specification. Along with the log-linear target portfolio this implies that (9) can be simply rewritten as

$$\delta_{i,t}(n) = \underbrace{\rho_{i,t} q_{i,t-1}(n)}_{\text{Partial Adjust.}} + \alpha_{i,t} p_t(n) + \epsilon_{i,t}(n) \tag{14}$$

where $\alpha_{i,t} = (1-\rho_{i,t})\beta_{i,t} + \rho_{i,t}$.[14] By separating the passive portfolio into lagged ownership and market equity, this allows separately identifying inertia from short-run elasticities. Note, that this is the same cross-sectional specification as in KY (2019) except for the lagged ownership term $q_{i,t-1}(n)$. The reduced-form coefficients $\alpha_{i,t}$ and $\rho_{i,t}$ allow backing out the investor's long and short-term elasticity as

$$\zeta_{i,t}^{SR} \approx 1 - \alpha_{i,t} \tag{15}$$

$$\zeta_{i,t}^{LR} \approx (1 - \rho_{i,t})^{-1}(1 - \alpha_{i,t}) \tag{16}$$

Between the instantaneous and long-term impact are h-quarter elasticities given by $\zeta_{i,t}^{(h)} \approx \sum_{s=0}^{h} \rho_{i,t}^{s}(1 - \alpha_{i,t})$. To rule out upward-sloping demand curves, I follow KY (2019) and impose the coefficient constraint $\alpha_{i,t} < 1$ for all investor-quarter pairs. I estimate both the partial adjustment model and the static logit specification (i.e. excluding $\rho_{i,t} q_{i,t-1}(n)$), quarterly from 1990 to 2024 for every investor using the same stock universes and pooling of narrow holdings across the estimations. To maximize comparability to the static

[14] The error term contains other demand shifters and is given by $\epsilon_{i,t}(n) = (1 - \rho_{i,t}) d_{i,t}(n) - \rho_{i,t} h_{i,t}$.

estimation, I use KY's mandate-driven demand to instrument for endogenous price levels in (14). Under valid instruments for market equity, we can consistently estimate the short-run elasticity $1 - \alpha_{i,t}$. However, extrapolating from the short to the long run by dividing by $(1 - \rho_{i,t})$ is only valid if $\rho_{i,t}$ is well identified.

C Identifying Inertia

If unobservable demand shocks are persistent, then $\rho_{i,t}$ is not consistently estimated. Intuitively, a long-run value investor may hold the same portfolio in the current quarter as last quarter not because she is inert, but because her value signals have not changed. However, note, that $q_{i,t-1}(n)$ are not past portfolio weights but past ownership *relative* to shares outstanding, i.e., total supply. The endogeneity concern is therefore less severe than it may initially appear. Persistent demand shocks affect portfolio weights but not necessarily relative ownership. I nevertheless provide two possible instruments for $q_{i,t-1}$ that are plausibly exogenous to current latent demand. First, I instrument $q_{i,t-1}$ by its long-run (3-year) lag. Since trading signals are unlikely to persist over several years, this helps mitigate the problem of omitted persistent factors that jointly affect past portfolios and current latent demand. Second, I use a model-based instrument by estimating the counterfactual lagged portfolio held in the absence of latent demand shocks, i.e., the portfolios implied by lagged observables. Because the results are largely unaffected by the choice of the instrument, the main text reports the results using raw lagged ownership $q_{i,t-1}(n)$. Appendix B.6 and B.7 provide further details and show that all results are qualitatively and quantitatively robust to the instrument choice. Last, in all specifications I impose the constraint that $\rho_{i,t} \in [0, 1)$. While for the majority of investors, the unconstrained $\hat{\rho}_{i,t}$ already lies between 0 and 1, this ensures finite long-run elasticities and unique equilibria.

D Estimation Results

Table 2 reports the estimated coefficients from the partial adjustment model across investor groups relative to the benchmark logit estimates without the partial adjustment term.

Portfolio Inertia. The first column reports inertia $\rho_{i,t}$ across investor groups. The average investor is relatively inert and partially trades towards their long-term desired portfolio with a speed of $1 - \rho_{i,t} = 0.16$. Bacchetta et al. (2023) find similarly low

adjustment speeds for international mutual funds using an alternative econometric specification with different instruments. The overall high inertia in institutional portfolios is consistent with the reduced form evidence on increasing rebalancing activity at longer horizons. Gradual portfolio adjustments lead to a large impact of demand shocks and excess volatility at short horizons that only slowly dissipate over time. Intuitively, changes in large positions should take longer to execute as trading greater fractions of current volume implies higher price impacts. In line with higher trading costs for bigger positions, I find that large investors are on average 23% more inert than small investors. The bottom rows of Table 2 confirm that results are robust to using different instruments for lagged portfolios, with adjustment speeds of 0.23 and 0.15 using long-run lags and the model-based instrument, respectively.

Short-run Elasticities. The second column reports the short-run elasticity $\zeta_{i,t}^{SR}$ from the partial adjustment model. The AUM-weighted average across all investors is 0.32. Among investors, elasticities exhibit an intuitive pattern: small investors are more elastic than large investors, value investors are more elastic than growth investors, and investors with high active shares (Cremers and Petajisto (2009)) are more elastic than those with low active shares. Large investors with a low active shares are the least elastic investor group with an average short-run elasticity of 0.12. Across all groups, short-run elasticities are lower than the static logit estimates.[15] Note that both estimators use the same instrument for endogenous market equities. This suggests that – conditional on the instrument – logit elasticities overstate how aggressively investors trade *quarterly* and instead capture what investors trade towards over longer horizons. Intuitively, an equal-weighted investor that rebalances their portfolio every year has a logit elasticity of around 1, but a quarterly short-run elasticity of 0 in most quarters.

Long-run Elasticities. Slow portfolio adjustments lead to implied long-run elasticities $\zeta_{i,t}^{LR} = \frac{\zeta_{i,t}^{SR}}{1-\rho_{i,t}}$ that are around 4 times higher than their short-term counterpart. The AUM-weighted average long-run elasticity implied by the partial adjustment model is 1.2. While many large investors are not very elastic at short horizons, they provide considerable liquidity at long horizons. Over the long run, large value investors are the most important liquidity providers with a long-run elasticity of 5.2. This emphasizes the importance of value investors in stabilizing equity markets over the long term. The static logit

[15] The slight difference to the average coefficient reported in Table 1, which use panel regressions and a pooled first stage to facilitate comparison with the reduced-form estimates.

elasticity lies between the short and the long-run elasticity from the partial adjustment model. This implies that the impact of counterfactuals based on logit demand systems may *underestimate* the true price response in the short run and *overestimate* the long-run effect. Note, that the average long-term elasticity implied by the partial adjustment model is slightly lower than the reduced-form long-run elasticity of 1.68 in Table 1. Naturally, the overall magnitudes of elasticities depend on the price instrument used. Less persistent price instruments such as flow-driven trades correspond to higher elasticities. However, the *ratios* between long and short-term elasticities are quantitatively consistent. Both, reduced-form and structural estimates suggest that long-term elasticities are 3 to 4 times larger than quarterly elasticities. This has important implications for the price impact of counterfactual experiments, which I explore in Section 8.

[Table 2 about here.]

E Decomposing the Difference in Dynamic versus Static Elasticities

What drives the difference between long-run elasticities from the partial adjustment model $\zeta_{i,t}^{LR}$ and the elasticities from static logit demand $\zeta_{i,t}^{KY}$? On the one hand, controlling for lagged holdings mitigates the problem of omitted persistent demand shocks that are correlated with long-run price levels. On the other hand, the partial adjustment model introduces heterogeneous trading horizons which may partly drive the divergence of $\zeta_{i,t}^{LR}$ and $\zeta_{i,t}^{KY}$. Appendix C illustrates the two effects in a simulated laboratory environment. Figure 2 shows that the difference in coefficients $\zeta_{i,t}^{LR} - \zeta_{i,t}^{KY}$ is strongly related to investor size. Large investors are considerably more elastic in the long run than what their static logit estimate implies. Note that because $\zeta_{i,t}^{LR} = \frac{\zeta_{i,t}^{SR}}{1-\rho_{i,t}}$ any difference between the two estimates can either be attributed to trading horizons ($\rho_{i,t} \neq 0$) or omitted variables ($\zeta_{i,t}^{SR} \neq \zeta_{i,t}^{KY}$). I therefore decompose the difference in coefficients into two components

$$\zeta_{i,t}^{LR} - \zeta_{i,t}^{KY} = TH_{i,t} + OV_{i,t} \tag{17}$$

where $TH_{i,t} = \zeta_{i,t}^{LR} - \zeta_{i,t}^{SR}$ captures variation due to trading horizons and $OV_{i,t} = \zeta_{i,t}^{SR} - \zeta_{i,t}^{KY}$ captures variation from omitted variables. In order to see why $OV_{i,t}$ captures omitted variables, note that the partial adjustment model (14) can be rewritten as

$$\tilde{\delta}_{i,t}(n) = \alpha_{i,t} p_t(n) + u_{i,t}(n).$$

where $\tilde{\delta}_{i,t} \equiv \delta_{i,t} - \rho_{i,t} q_{i,t-1}$ are quasi-holdings corrected for persistent demand shifts. Therefore any difference between $\zeta_{i,t}^{SR}$ and $\zeta_{i,t}^{KY}$ is driven by correcting the left-hand side for persistent demand shifts that are correlated with prices. The grey and red bars in Figure 2 plot the decomposition of $\zeta_{i,t}^{LR} - \zeta_{i,t}^{KY}$ into $TH_{i,t}$ and $OV_{i,t}$. The difference in elasticities is primarily driven by trading horizons, i.e. by inertia $\rho_{i,t}$. A simple variance decomposition reveals that 84% (std. error 1.4%) of the variation in $\zeta_{i,t}^{LR} - \zeta_{i,t}^{KY}$ is driven by $TH_{i,t}$.[16] While the omitted variable effect is 0 for large investors, it is negative for small investors. The portfolios of smaller investors are tilted towards small-cap stocks, which gives the impression of elasticity when failing to control for past holdings. Therefore static logit elasticities overstate short-run trading activity for smaller institutions.

[Figure 2 about here.]

7 A New Measure of Long-Term Reversal

As implied by Proposition 1, the price multiplier matrices based on short and long-run elasticities are respectively given by the ownership-weighted sum of the investor-specific elasticities

$$\mathcal{M}_t^{SR} \equiv \left(\sum_{i=1}^{I} \text{diag}(Q_{i,t}) \zeta_{i,t}^{SR} \right)^{-1} \tag{18}$$

$$\mathcal{M}_t^{LR} \equiv \left(\sum_{i=1}^{I} \text{diag}(Q_{i,t}) \zeta_{i,t}^{LR} \right)^{-1}. \tag{19}$$

Because investors are inert, elasticities are larger in the long run and the price impact from demand shocks reverts over time. I define long-term reversal as the difference in long- and short-run multipliers

$$\mathcal{R}_t(n) \equiv \mathcal{M}_t^{SR}(n) - \mathcal{M}_t^{LR}(n). \tag{20}$$

which is a novel stock-level measure of reversal. Because $\rho_{i,t} \in [0, 1)$ and $\alpha_{i,t} < 1$, short-run price impact (ceteris paribus) always exceeds long-run price impact and $\mathcal{R}_t(n) \geq 0$. Two factors drive reversal. First, stocks in the habitat of short-term investors (with low $\rho_{i,t}$) have similar long and short-term impacts and therefore exhibit less reversal. Second,

[16] Formally, the variance decompositioin is given by $1 = \frac{\text{cov}\left(\zeta_{i,t}^{LR} - \zeta_{i,t}^{KY}, TH_{i,t}\right)}{\text{var}\left(\zeta_{i,t}^{LR} - \zeta_{i,t}^{KY}\right)} + \frac{cov(\zeta_{i,t}^{LR} - \zeta_{i,t}^{KY}, OVB_{i,t})}{var(\zeta_{i,t}^{LR} - \zeta_{i,t}^{KY})}$.

stocks in the habitat of long-term investors (with high $\rho_{i,t}$) are particularly exposed to reversal if these investors have high elasticities. This encapsulates the difference in short and long-run dynamics which is critically driven by the interaction of inertia and elasticity. An intuitive example is stocks in the habitat of long-run value investors. The initial price impact may considerably exceed the long-run impact if the primary investors in the stock become more elastic over time.

Panel a) of Figure 3 plots the stock-specific multipliers and reversal implied by the dynamic demand system. The average short-run multiplier across stocks is 4.5 and strongly declines over time to around 1. This suggests that the impact of counterfactual experiments is over 4 times larger in the short run than over long horizons. Again, while the overall magnitude of counterfactual price changes likely depends on the instrument with which elasticities are identified, Figure 3 suggests that there are important differences in the *relative* impact over time. Note, that while investor-specific elasticities do not vary across stocks (assuming $w_{i,t}(n) \approx 0$), ownership shares $Q_{i,t}(n)$ vary considerably. This implies cross-sectional variation in multipliers across stocks. Panel b) plots the distribution of the dynamic multiplier. There is considerable heterogeneity in price impact and reversal in the cross-section of stocks. Some stocks are more inert than others which results in heterogeneous reversal speeds. The next section explores the return predictability from this novel measure of stock-specific long-term reversal.

[Figure 3 about here.]

8 Applications

A Reversal of Non-fundamental Price Pressure

Coval and Stafford (2007) and Lou (2012) find evidence of significant price reversal following flow-induced trades by mutual funds. Stocks with higher $\mathcal{R}_t(n)$ should exhibit greater reversal following non-fundamental demand shifts. Note, that stock-specific reversal is obtained quarterly without forward-looking information. At each quarter, I sort stocks into high and low reversal stocks, based on $\mathcal{R}_t(n)$ being above and below the cross-sectional median respectively. Within the high and low reversal stocks, I form in- and outflow portfolios based on flow-induced trades. Specifically, I compute the value-weighted average return of stocks in the top (bottom) tercile of flow-induced traders. According to the model, the initial dispersion in returns of the in- and outflow portfolios should converge more strongly over time for reversal stocks than for non-reversal stocks.

Similarly, a long-short portfolio that buys outflow stocks and sells inflow stocks should achieve higher returns when conditioning on high reversal stocks. Figure 4 confirms these predictions. Among low reversal stocks, the price impact of mutual fund flows does not exhibit significant return reversal. While standard errors widen considerably over time, cumulative long-short returns remain significantly positive for at least 7 quarters after the initial flow. Instead for high reversal stocks, the price impact fully reverts within 3 quarters. This highlights the cross-sectional information contained in the reversal measure beyond its overall magnitude. As reversal is generally difficult to estimate from long-term returns (due to high volatility at long horizons), the simple holdings-based measure is a useful alternative. Many trading strategies exploit the price impact (alpha) arising from non-fundamental demand shocks of different investors. Future research could condition the turnover of such trading strategies on the speed of alpha decay implied by $\mathcal{R}_t(n)$.

[Figure 4 about here.]

B Simulating Dynamic Price Paths from Counterfactual Experiments

The counterfactual prices obtained from the demand system depend on the multiplier matrix implied by the elasticities (see Proposition 1). A simple out-of-sample test of the demand system is whether it *quantitatively* predicts the price path following exogenous demand shocks. Ideally, these demand shocks have not been used for identifying the system. The elasticities from the partial adjustment model have been identified from the mandate-based instrument. Therefore the impact of mutual funds' flow-induced trades provides a true out-of-sample test of the estimated elasticities and the long-term reversal implied by the demand system. Under the assumption, that flow-induced trades are *exogenous* demand shifts, one can test the demand system by comparing the simulated price impact to the observed impact. The grey line in Figure 5 plots the cumulative impact of flow-induced trades over time.[17] I then simulate the price impact of the same flow-induced trades using the horizon-specific multiplier matrices from the partial adjustment model. I also plot the static price path implied by the standard logit demand system without partial adjustment. The counterfactual impact implied by the partial adjustment model impressively matches the observed impact and its reversal both qualitatively and quantitatively. Instead, static logit demand implies a constant price impact that underestimates the short-run and overestimates the long-run effect.

[17] I estimate the cumulative future price impact following a one time demand shock at $t-h$ using a distributed lag model given by $\Delta p_t(n) = \sum_{l=1}^{h} \beta_l \Delta z_{i,t-l}(n) + u_{i,t}(n)$ including time-fixed effects.

[Figure 5 about here.]

C Reversal and Return Predictability

I test whether $\mathcal{R}_t(n)$, a new measure of long-term reversal, contains additional information beyond standard return predictors. Reversal should only affect returns after a demand shock. If momentum is partly driven by non-fundamental shocks, then stocks with higher $\mathcal{R}_t(n)$ should exhibit weaker momentum: As investors become more elastic over time, they trade more aggressively against the initial demand shocks, thereby acting contrarian against momentum. I test this by conditioning a momentum strategy on reversal. To this end, I construct a standard momentum portfolio that buys (sells) the top (bottom) tercile of stocks based on past cumulative 12-1 month returns. At each formation date, I split the sample into two subgroups — one with stocks whose reversal $\mathcal{R}_t(n)$ is above the cross-sectional median and one below — and then construct tercile momentum portfolios for each subgroup. Figure 6 shows that among reversal stocks – those held by increasingly elastic investors, such as long-run value investors – momentum disappears entirely. In contrast, non-reversal stocks exhibit over 3% higher annualized momentum returns, with the Sharpe ratio of the unconditional strategy rising from 0.37 to 0.54.

[Figure 6 about here.]

Table 3 confirms this using monthly Fama-MacBeth regressions. The baseline coefficient on the unconditional momentum characteristic is 0.26, meaning that expected monthly returns increase by 0.26% for a one standard deviation increase in momentum. I then condition on reversal by interacting momentum with a dummy variable equal to 1 if a stock has above median reversal in the prior month. The coefficient on the interaction is -0.5%, while the unconditional coefficient doubles to 0.47%. Hence, momentum returns disappear entirely for reversal stock and are twice as high for non-reversal stocks. This underscores the incremental information embedded in the stock-specific reversal measure and the broader role of portfolio rebalancing dynamics in shaping return predictability.

[Table 3 about here.]

9 Conclusion

A growing literature uses portfolio holdings to quantify the impact of demand shocks on equilibrium prices via investors' price elasticity of demand. Existing studies estimate

demand slopes using static logit regressions, implying constant price impacts in counterfactual experiments. In this paper, I provide both reduced-form and structural evidence that investors' price elasticity of demand increases substantially at longer horizons, reaching up to four times its short-term value. This suggests a key distinction between what investors hold in the cross-section and how they trade in the time series. The price impact of counterfactual demand shocks is therefore larger at high frequencies but dissipates over time as investors adjust their portfolios. To capture these dynamics, I develop a simple yet flexible extension of the logit framework – a partial adjustment model – that links holdings and trades within a unified demand system. I show that quadratic transaction costs are a possible micro-foundation for this model, rationalizing investors' gradual portfolio rebalancing. The dynamic model accurately predicts both the short-run impact of demand shocks and their long-run reversal out-of-sample. These findings emphasize the importance of accounting for trading horizons and portfolio adjustment speeds into demand-based asset pricing models, ensuring more accurate counterfactual predictions.

References

Angrist, J. D. and A. B. Krueger (1992). The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American statistical Association 87*(418), 328–336.

Bacchetta, P., S. Tieche, and E. Van Wincoop (2023). International portfolio choice with frictions: Evidence from mutual funds. *The Review of Financial Studies 36*(10), 4233–4270.

Balasubramaniam, V., J. Y. Campbell, T. Ramadorai, and B. Ranish (2021). Who owns what? a factor model for direct stock holding. Technical report, National Bureau of Economic Research.

Barber, B. M. and T. Odean (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance 55*(2), 773–806.

Benetton, M. and G. Compiani (2020). Investors' beliefs and asset prices: A structural model of cryptocurrency demand. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2020-107).

Betermier, S., L. E. Calvet, S. Knüpfer, and J. Kvaerner (2022). What do the portfolios of individual investors reveal about the cross-section of equity returns? *Available at SSRN 3795690*.

Bogousslavsky, V. (2016). Infrequent rebalancing, return autocorrelation, and seasonality. *The Journal of Finance 71*(6), 2967–3006.

Bogousslavsky, V., P. Collin-Dufresne, and M. Sağlam (2021). Slow-moving capital and execution costs: Evidence from a major trading glitch. *Journal of Financial Economics 139*(3), 922–949.

Bohn, H. and L. L. Tesar (1996). Us equity investment in foreign markets: portfolio rebalancing or return chasing? *The American Economic Review 86*(2), 77–81.

Bretscher, L., L. Schmid, I. Sen, and V. Sharma (2024). Institutional corporate bond pricing. *Swiss Finance Institute Research Paper* (21-07).

Calvet, L. E., J. Y. Campbell, and P. Sodini (2009). Fight or flight? portfolio rebalancing by individual investors. *The Quarterly journal of economics 124*(1), 301–348.

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*(1), 212–246.

Chaudhary, M., Z. Fu, and J. Li (2023). Corporate bond multipliers: Substitutes matter. *Available at SSRN*.

Collin-Dufresne, P., K. Daniel, and M. Sağlam (2020). Liquidity regimes and optimal dynamic asset allocation. *Journal of Financial Economics 136*(2), 379–406.

Coqueret, G. (2022). Characteristics-driven returns in equilibrium. *arXiv preprint arXiv:2203.07865*.

Coval, J. and E. Stafford (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics 86*(2), 479–512.

Cremers, K. M. and A. Petajisto (2009). How active is your fund manager? a new measure that predicts performance. *The review of financial studies 22*(9), 3329–3365.

Darmouni, O., K. Siani, and K. Xiao (2022). Nonbank fragility in credit markets: Evidence from a two-layer asset demand system. *Available at SSRN 4288695*.

Duffie, D. (2010). Presidential address: Asset price dynamics with slow-moving capital. *The Journal of finance 65*(4), 1237–1267.

Edmans, A., I. Goldstein, and W. Jiang (2012). The real effects of financial markets: The impact of prices on takeovers. *The Journal of Finance 67*(3), 933–971.

Enke, B. and T. Graeber (2023). Cognitive uncertainty. *The Quarterly Journal of Economics 138*(4), 2021–2067.

Fama, E. F. and K. R. French (2002). Testing trade-off and pecking order predictions about dividends and debt. *Review of financial studies*, 1–33.

Fang, X., B. Hardy, and K. K. Lewis (2022). Who holds sovereign debt and why it matters. Technical report, National Bureau of Economic Research.

Froot, K. A., P. G. O'connell, and M. S. Seasholes (2001). The portfolio flows of international investors. *Journal of financial Economics 59*(2), 151–193.

Gabaix, X. and R. S. Koijen (2023). In search of the origins of financial fluctuations: The inelastic markets hypothesis. *Available at SSRN 3686935*.

Gabaix, X., R. S. Koijen, F. Mainardi, S. Oh, and M. Yogo (2024a). Asset demand of us households. Technical report, National Bureau of Economic Research.

Gabaix, X., R. S. Koijen, F. Mainardi, S. Oh, and M. Yogo (2024b). Limited risk transfer between investors: A new benchmark for macro-finance models. *Available at SSRN*.

Gârleanu, N. and L. H. Pedersen (2013). Dynamic trading with predictable returns and transaction costs. *The Journal of Finance 68*(6), 2309–2340.

Gârleanu, N. and L. H. Pedersen (2016). Dynamic portfolio choice with frictions. *Journal of Economic Theory 165*, 487–516.

Greenwood, R. (2005). Short-and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage. *Journal of Financial Economics 75*(3), 607–649.

Grossman, S. J. and M. H. Miller (1988). Liquidity and market structure. *the Journal of Finance 43*(3), 617–633.

Haddad, V., P. Huebner, and E. Loualiche (2021). How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing. *Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing (April 7, 2021)*.

Han, X., N. L. Roussanov, and H. Ruan (2021). Mutual fund risk shifting and risk anomalies. *Available at SSRN 3931449*.

Hartzmark, S. M. and D. H. Solomon (2021). Predictable price pressure. *Available at SSRN 3853096*.

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies 28*(3), 650–705.

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

Jansen, K. A. (2024). Long-term investors, demand shifts, and yields. *The Review of Financial Studies*, hhae071.

Jansen, K. A., W. Li, and L. Schmid (2024). Granular treasury demand with arbitrageurs. Technical report, National Bureau of Economic Research.

Jiang, Z., R. J. Richmond, and T. Zhang (2022). Understanding the strength of the dollar. Technical report, National Bureau of Economic Research.

Jiang, Z., R. J. Richmond, and T. Zhang (2024). A portfolio approach to global imbalances. *The Journal of Finance 79*(3), 2025–2076.

Kashyap, A. K., N. Kovrijnykh, J. Li, and A. Pavlova (2023). Is there too much benchmarking in asset management? *American Economic Review 113*(4), 1112–1141.

Kayhan, A. and S. Titman (2007). Firms' histories and their capital structures. *Journal of financial Economics 83*(1), 1–32.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Koijen, R. S., F. Koulischer, B. Nguyen, and M. Yogo (2021). Inspecting the mechanism of quantitative easing in the euro area. *Journal of Financial Economics 140*(1), 1–20.

Koijen, R. S., R. J. Richmond, and M. Yogo (2024). Which investors matter for equity valuations and expected returns? *Review of Economic Studies 91*(4), 2387–2424.

Koijen, R. S. and M. Yogo (2019). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

Koijen, R. S. and M. Yogo (2020). Exchange rates and asset prices in a global demand system. Technical report, National Bureau of Economic Research.

Lettau, M., S. C. Ludvigson, and P. Manoel (2018). Characteristics of mutual fund portfolios: Where are the value funds? Technical report, National Bureau of Economic Research.

Li, J. and Z. Lin (2022). Price multipliers are larger at more aggregate levels. *Available at SSRN 4038664*.

Lou, D. (2012). A flow-based explanation for return predictability. *The Review of Financial Studies 25*(12), 3457–3489.

Mitchell, M., L. H. Pedersen, and T. Pulvino (2007). Slow moving capital. *American Economic Review 97*(2), 215–220.

Pavlova, A. and T. Sikorskaya (2022). Benchmarking intensity.

Petajisto, A. (2009). Why do demand curves for stocks slope down? *Journal of Financial and Quantitative Analysis 44*(5), 1013–1044.

Schmickler, S. (2020). Identifying the price impact of fire sales using high-frequency surprise mutual fund flows. *Available at SSRN 3488791*.

Shleifer, A. (1986). Do demand curves for stocks slope down? *The Journal of Finance 41*(3), 579–590.

Tamoni, A., S. Sokolinski, and Y. Li (2024). Which investors drive anomaly returns and how? *Available at SSRN 4242745*.

Wardlaw, M. (2020). Measuring mutual fund flow pressure as shock to stock returns. *The Journal of Finance 75*(6), 3221–3243.

Wurgler, J. and E. Zhuravskaya (2002). Does arbitrage flatten demand curves for stocks? *The Journal of Business 75*(4), 583–608.

Figure 1: **How much variation in portfolios is explained by lagged holdings?**
The figure plots R^2s from simple univariate cross-sectional regressions of investor-specific portfolio weights onto a passive buy-and-hold portfolios $w^p_{i,t}(n) = \frac{Q_{i,t-h}(n)P_t(n)}{\sum_n Q_{i,t-h}(n)P_t(n)}$ with h ranging from 1 to 8 quarters. For each institution-quarter pair, I estimate $w(n) = \alpha + \beta w^p(n) + \epsilon(n)$ and report the distribution and average R^2 (dotted lines) for each horizon.

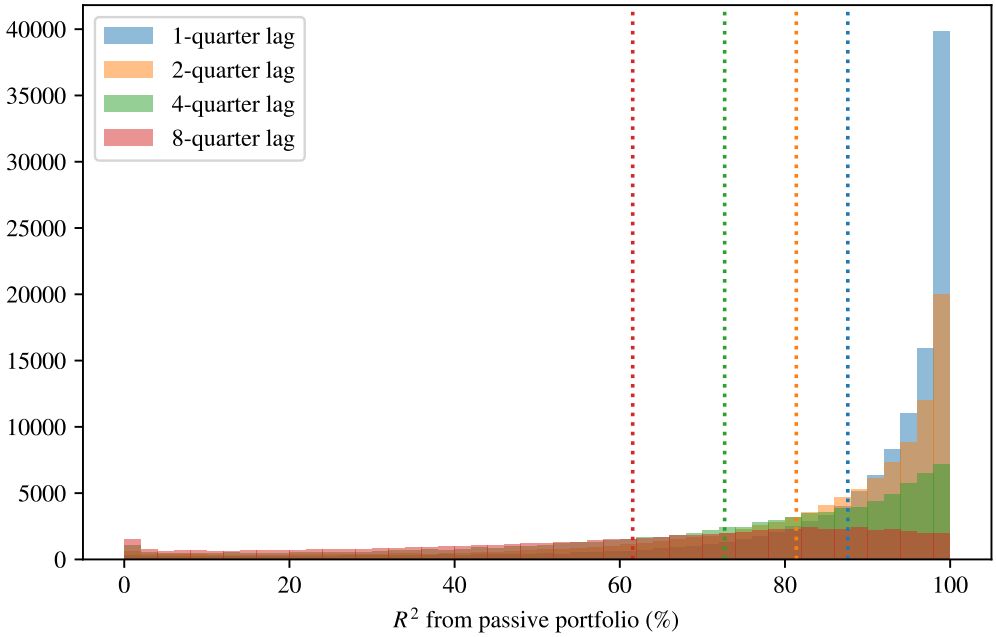


Figure 2: **Long-run Elasticities versus Logit Elasticities**.
The figure decomposes the difference between long-run elasticities from the partial adjustment model $\zeta_{i,t}^{LR}$ and the static elasticities from logit demand $\zeta_{i,t}^{KY}$. The black bars plot $\zeta_{i,t}^{LR} - \zeta_{i,t}^{KY}$, averaged across AUM deciles. The red and grey bars decompose the difference into heterogeneity in trading horizons $TH_{i,t} = \zeta_{i,t}^{LR} - \zeta_{i,t}^{SR}$ and omitted variables $OV_{i,t} = \zeta_{i,t}^{SR} - \zeta_{i,t}^{KY}$.

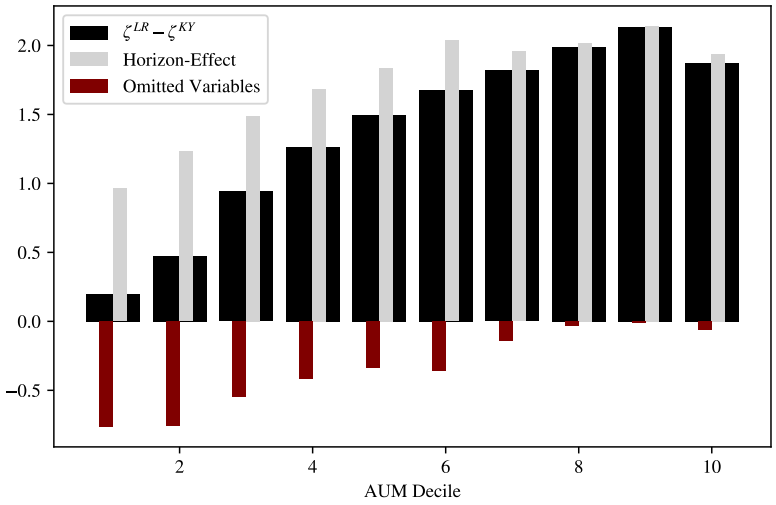
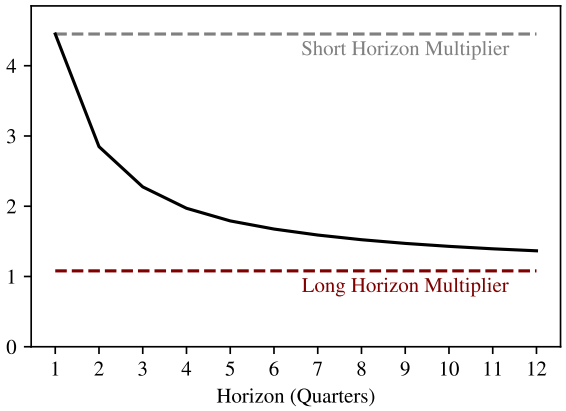


Figure 3: **Stock-specific Dynamic Multipliers**.
Panel a) of the figure plots the multiplier (i.e., price impact of counterfactual experiments) over different horizons $\mathcal{M}_t^h \equiv \left(\sum_{i=1}^{I} \text{diag}(Q_{i,t})\zeta_{i,t}^{(h)}\right)^{-1}$. I report the median $\mathcal{M}_t^h$ across stocks and quarters for horizons ranging from 1 to 12 quarters. Panel b) plots the distribution of the multiplier across stocks.

(a) Short versus Long-run Impact



(b) Distribution of Multiplier

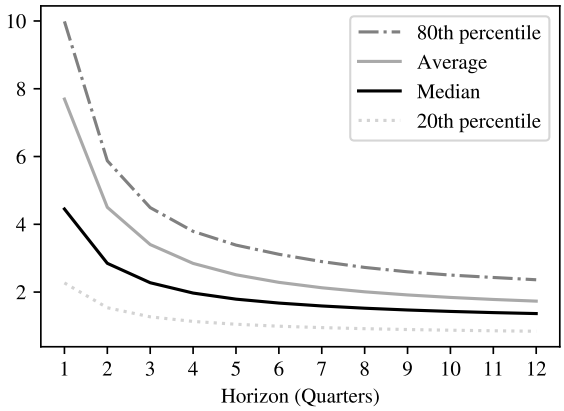


Figure 4: **Explaining the Reversal of Flow-Induced Trades**.
The figure compares the cumulative flow-induced price pressure for high versus low reversal stocks. At each quarter, I split the cross-section of US stocks into low (left panel) and high (right panel) reversal stocks based on $\mathcal{R}_t(n)$ being above and below the cross-sectional median. Within each sample, I sort stocks into value-weighted portfolios based on top and bottom tercile flow-driven trades. I report average cumulative returns in excess of the CRSP value-weighted market return. The bottom panels report the average long-short portfolio return of in- versus outflow stocks. The shaded areas indicate 95% confidence intervals.

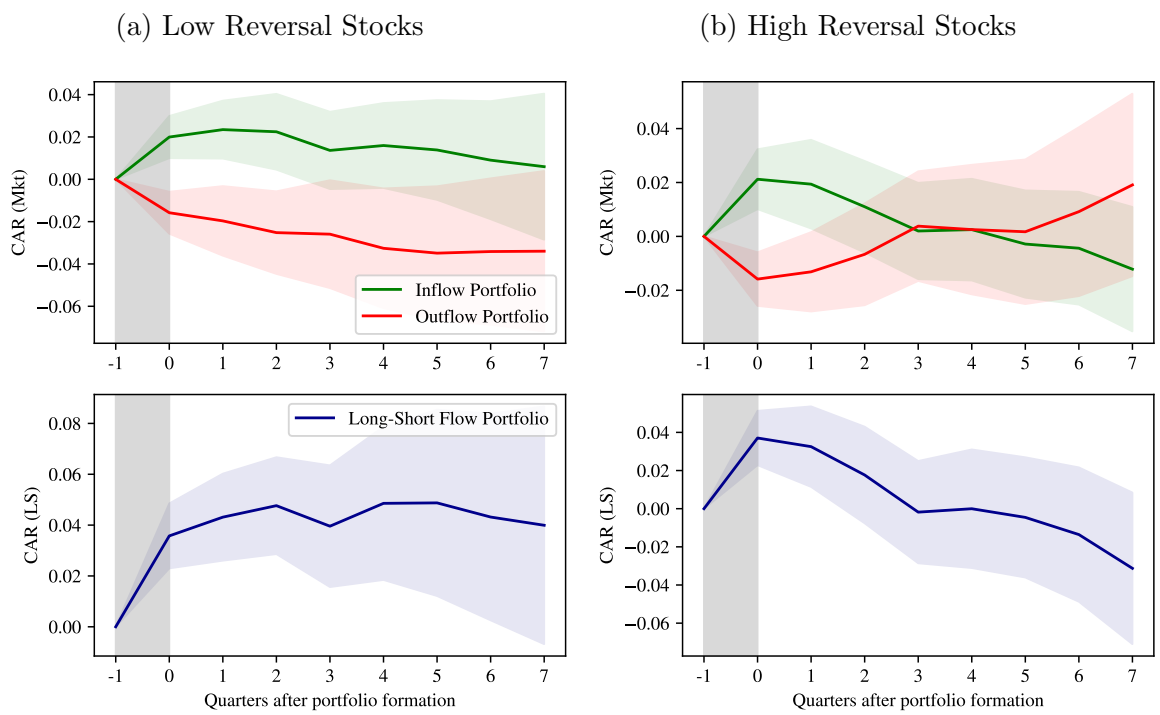


Figure 5: **Counterfactual Experiments Out-of-Sample**.
The figure compares the price impact of flow-induced trades implied by the static logit demand system to the price impact implied by the partial adjustment demand system. I compare the model-implied price impact to the *true* impact estimated using a distributed lag model $\Delta p_t(n) = \sum_{l=1}^{h} \beta_l \Delta z_{i,t-l}(n) + u_{i,t}(n)$ including time-fixed effects and stock-specific control variables given by log book equity, market beta, profitability, investment and dividends-to-book equity. I plot the cumulative coefficients and 95% confidence intervals from a Wald test with standard errors clustered by time. The red and blue lines report the structurally implied impact, $\frac{1}{T}\sum_{t=1}^{T} \text{cov}_t\left(\mathcal{M}^h(n), \Delta z(n)\right) / \text{var}_t\left(\Delta z(n)\right)$, which is a weighted average of the stock-specific multiplier, where stock-time observations are weighted proportional to their importance in the OLS regressions.

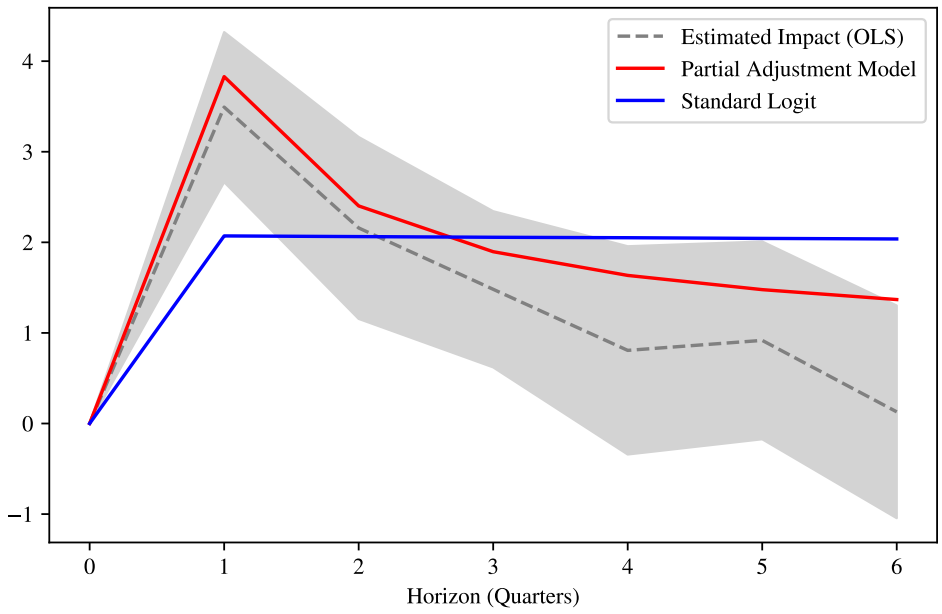


Figure 6: **Inertia and Momentum Returns**.
The figure plots the cumulative returns of monthly rebalanced momentum portfolios. Momentum is defined as the past cumulative returns from 12 months to 1 month prior to the portfolio formation date $[t-12, t-1]$. The black line is the return on a portfolio that goes long (short) the top (bottom) tercile of momentum stocks. Portfolios are equal-weighted in each leg. The blue and red lines report conditional momentum returns, by splitting the sample in two halves based on the median stock-specific reversal $\mathcal{R}_t(n)$ in the formation month and computing conditional momentum terciles.

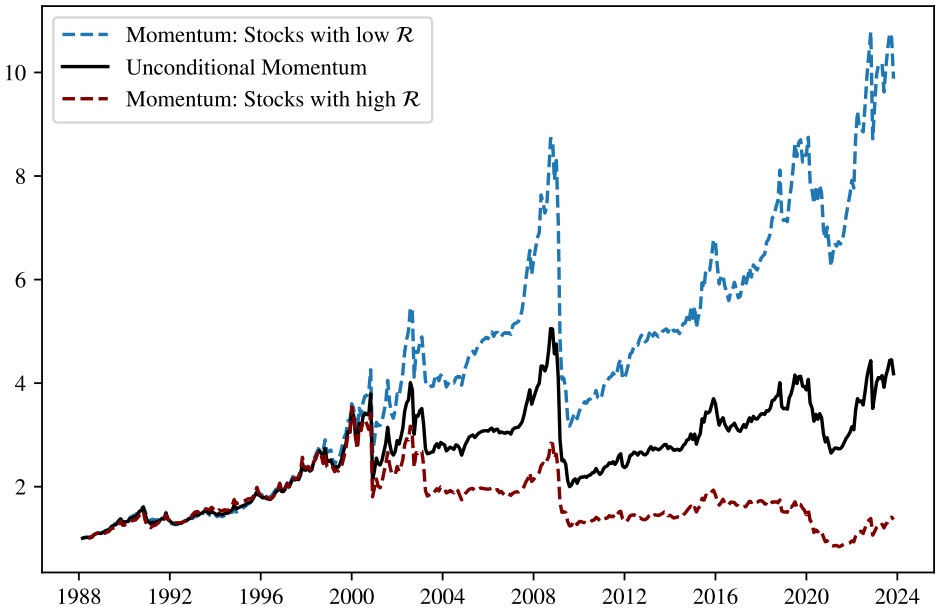


Table 1: **Reduced-Form Elasticities at Different Horizons.**
The table reports the estimated elasticities from the simple difference-in-difference estimator

$$\text{1st Stage:} \quad \Delta^h p_t(n) = m^{(h)} \Delta^h z_t(n) + u_t(n)$$
$$\text{2nd Stage:} \quad \Delta^h q_{i,t}(n) = -\zeta_i^{(h)} \Delta^h \hat{p}_t(n) + d_{i,t}(n)$$

for horizons ranging from $h = 1$ quarter up to 4 years. $\Delta^h z_t(n)$ denotes the flow-based instrument constructed over h periods. $d_{i,t}(n)$ and $u_t(n)$ include time-fixed effects and stock-specific control variables, log book equity, dividends to book equity, profitability, investment, and market beta. I follow KY (2019) and impose $\zeta_i^{(h)} > 0$ for all investors at all horizons. Column 1 reports the AUM-weighted average elasticity across all investors where AUM-weights are averaged over the sample period. The standard error of the AUM-weighted average elasticity is computed by bootstrapping 100 draws from the standard errors on the investor-specific elasticities and constructing the AUM-weighted average for each sample. Columns 2-4 report the distribution of elasticities across all investors, excluding investors with binding constraints ($\zeta_i^{(h)} = 0$). The rightmost column reports the first-stage coefficient, $m^{(h)}$, estimated from a pooled regression across all stocks and time periods. Standard errors are clustered by time. The bottom row reports the static logit elasticity identified from KY's mandate-based instrument. Across all specifications, I use the same binning across investors, panel estimation, and pooled first stage.

Horizon h (Quarters)	Avg. Elasticity ζ_i^h		Elasticity Distribution			First Stage	
			10th pctl.	Median	90th pctl.		
1	0.58	(0.2)	0.14	0.47	0.8	2.57	(0.38)
2	0.68	(0.21)	0.35	0.61	1.01	2.2	(0.3)
3	0.85	(0.22)	0.24	0.73	1.25	1.94	(0.25)
4	0.72	(0.16)	0.36	0.8	1.39	1.71	(0.23)
5	0.81	(0.17)	0.25	0.92	1.55	1.53	(0.19)
6	0.87	(0.18)	0.3	1.07	1.88	1.37	(0.15)
7	0.99	(0.15)	0.24	1.07	2.21	1.23	(0.12)
8	1.0	(0.11)	0.42	1.2	2.41	1.15	(0.11)
9	1.02	(0.12)	0.38	1.21	2.6	1.06	(0.11)
10	1.14	(0.13)	0.42	1.22	2.52	0.97	(0.13)
11	1.35	(0.14)	0.47	1.53	3.17	0.89	(0.14)
12	1.41	(0.13)	0.5	1.78	3.86	0.82	(0.15)
13	1.43	(0.13)	0.51	1.91	4.34	0.77	(0.16)
14	1.55	(0.15)	0.3	1.92	3.94	0.72	(0.16)
15	1.59	(0.16)	0.67	2.19	4.67	0.66	(0.17)
16	1.68	(0.17)	0.87	2.5	5.02	0.61	(0.18)
Static Logit (Levels)	0.73	(0.01)	1.52	2.18	2.78	0.58	(0.02)

Table 2: **Structural Elasticities from Partial Adjustment Model.**
The table reports the results of estimating the constrained dynamic demand curve $\delta_{i,t}(n) = \rho_{i,t} q_{i,t-1}(n) + \alpha_{i,t} p_t(n) + X_t \beta_{i,t} + \epsilon_{i,t}(n)$ from 1990 to 2024 quarterly for every investor. The moment condition is given by

$$\mathbb{E}_t[\epsilon_{i,t}(n) | q_{i,t-1}(n), z^M_{i,t}(n), X_t] \quad \text{s.t.} \quad \rho_{i,t} \in [0,1), \alpha_{i,t} < 1 \quad \forall i,t.$$

where X_t is a vector of controls including a constant, log book equity, market beta, dividends to book equity, investment, and profitability, and $z^M_{i,t}(n)$ is KY's mandate-based instrument. Short and long-run elasticities are given by $\zeta^{SR}_{i,t} = 1 - \alpha_{i,t}$ and $\zeta^{LR}_{i,t} = (1 - \rho_{i,t})^{-1}(1 - \alpha_{i,t})$ respectively. I group investors into deciles by size, active share, and value/growth tilts and report average coefficients for each group. Active shares are computed as $\frac{1}{2}\sum_n |w_{i,t}(n) - w^M_{i,t}(n)|$ where $w^M_{i,t}(n)$ are value-weights within the investor's subset of stocks held. Value/growth tilts are computed as portfolio-weighted averages of market-to-book ratios following Lettau et al. (2018). The bottom rows report the AUM-weighted average elasticity across investors when instrumenting $q_{i,t-1}(n)$ with its 3-year lag and the model-based instrument respectively. The rightmost column reports the elasticities obtained from a static quarterly logit estimation following KY (2019).

	Dynamic Estimates from Partial Adjustment Model			
	Inertia $\rho_{i,t}$	Short-Run Elasticity $\zeta^{SR}_{i,t}$	Long-Run Elasticity $\zeta^{LR}_{i,t} = \zeta^{SR}_{i,t}/(1-\rho_{i,t})$	Logit Elasticity $\zeta^{KY}_{i,t}$
All Investors (AUM-weighted)	0.84	0.31	1.19	0.5
Large Investors	0.8	0.73	2.66	0.79
Active	0.8	1.05	3.72	1.23
Passive	0.87	0.12	0.7	0.11
Growth	0.76	1.09	3.59	1.08
Value	0.75	1.58	5.18	1.69
Small & Med. Investors	0.65	0.91	2.61	1.28
Active	0.64	1.02	2.83	1.33
Passive	0.67	0.76	2.26	1.09
Growth	0.63	0.98	2.68	1.43
Value	0.61	1.14	2.97	1.5
Robustness to Instruments for Past Portfolios				
Long-run Lags	0.77	0.32	1.35	
Model-based Instrument	0.85	0.43	1.21	

Table 3: **Inertia and Momentum Returns: Fama-McBeth Regressions**
The table plots the coefficients to Fama-McBeth (FMB) regressions of monthly stock returns (in excess of the 1-month Treasury-bill rate) onto lagged stock-specific characteristics. The left column reports the baseline FMB regression. The right column interacts momentum with a dummy $1_{\mathcal{R}}$ equal to 1 if a stock has above median reversal $\mathcal{R}_{t-1}(n)$ in the prior month $t-1$.

Characteristics	Baseline		Conditional	
Log market equity	-0.27	(0.08)	-0.28	(0.08)
Book-to-market equity	0.08	(0.05)	0.08	(0.05)
Profitability	0.19	(0.05)	0.2	(0.05)
Investment	-0.36	(0.03)	-0.36	(0.03)
Market beta	0.08	(0.09)	0.08	(0.09)
Momentum	0.26	(0.09)	0.47	(0.1)
Momentum$\times 1_{\mathcal{R}}$			-0.52	(0.08)

Appendix A Proofs and Derivations

A.1 Intuition from a CARA Utility Investor

Assume that the stocks $n = 1, ..., N$ pay risky future cashflows $D(n)$ that are jointly normally distributed with a covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$. The investors' initial endowments $A_{i,0}$ are exogenously determined e.g. by the money that households allocate to them. The funds may invest in the N stock, as well as a riskless asset whose exogenous interest rate is normalized to 0. Under CARA preferences with risk aversion equal to γ_i, the institution maximizes $\mathbb{E}[e^{-\gamma_i A_i}]$, subject to the budget constraint on future wealth $A_i = A_{i,0} + Q_i'(\mathbb{E}[D] - P)$, by choosing the optimal portfolio

$$Q_i = (\gamma_i \Sigma)^{-1}(\mathbb{E}[D] - P). \tag{21}$$

Formally, the elasticity is given by $\text{diag}(Q_{i,t})^{-1}(\gamma_i \Sigma)^{-1}\text{diag}(P_t)$. For CARA preferences it is more convenient to define elasticities in absolute terms as opposed to percentages, in which case the elasticity matrix is given by $\tilde{\zeta}_i = -\frac{\partial Q_i}{\partial P'} = (\gamma_i \Sigma)^{-1}$. Market clearing implies that $\sum_i Q_i = Q^*$. The linear demand function implies that equilibrium market clearing prices are available in closed form and given by the CAPM

$$P = \mathbb{E}[D] - \gamma_i \Sigma Q^* \tag{22}$$

where $\gamma_i \equiv 1/\left(\sum_{i=1}^{I} 1/\gamma_i\right)$ is the market's effective risk aversion. In the CARA case, the most intuitive way to model exogenous demand ΔD is via a change in supply Q^*, such as the inclusion of some stocks in an index. The resulting price change is given by $\Delta P = \gamma_i \Sigma \Delta Q^*$. Thus the CARA setting implies that the multiplier matrix is proportional to $\gamma_i \Sigma$. It is driven by the covariance of cashflows and the market's effective risk aversion. Riskier stocks have a larger multiplier because they contribute more to the arbitrage portfolio. Note, that the channel through which supply shocks affect equilibrium prices in the CAPM is entirely risk-driven: If investors are risk neutral, or future cashflows are risk-free then $P = \mathbb{E}[D]$ and supply shocks do not affect prices. Greenwood (2005) shows that in a multi-period CARA setting the price impact of supply shocks decreases over time as uncertainty is resolved and the risk of future cashflows converges to 0. Formally, $\Delta P = \gamma_i \Sigma \Delta Q^*(T-t)$ where $T-t$ is the time left until liquidation time T. In other words, investors become more elastic over time as future payouts become more predictable. In the limit, two assets with equal riskless cashflows trade at identical prices.

A.2 Long-Run Elasticities

An investor's optimal dynamic portfolio is given by

$$Q_{i,t} = f_i\big(P_t, X_t, Q_{i,t-1}\big) \tag{23}$$

Via the lagged holdings term $Q_{i,t-1}$, the optimal portfolio depends on infinitely many lagged characteristics and prices. Section A.5 provides a possible microfoundation. The short-run elasticity captures the immediate response to price changes given by

$$\zeta_{i,t}^{SR} = -\text{diag}(Q_{i,t})^{-1}\frac{\partial Q_{i,t}}{\partial P_t'}\text{diag}(P_{i,t}). \tag{24}$$

The long-horizon elasticity accounts for the cumulative impact of a change in prices at time t, including both the direct impact at t and all indirect impacts via $Q_{i,t+1}, Q_{i,t+2}, \ldots$. The chain rule implies $\frac{\partial Q_{i,t+s}}{\partial P_t} = \frac{\partial f_i(P_{t+s}, X_{t+s}, Q_{i,t+s-1})}{\partial Q_{i,t+s-1}'}\frac{\partial Q_{i,t+s-1}}{\partial P_t'}$. For simplicity, I assume that inertia is time-independent and given by $\frac{\partial f_i}{\partial Q_{i,t+k-1}'} = \rho_i \mathbf{I}$ where ρ_i is a scalar.[18] The long-run response to a price change at time t is therefore

$$\frac{\partial Q_{i,t+s}}{\partial P_t'} = \frac{\partial Q_{i,t}}{\partial P_t'}\rho_i^s \tag{25}$$

While this gives the instantaneous response to a lagged price change, the long-horizon elasticity is defined as the total cumulative response as s goes to infinity

$$\zeta_{i,t}^{LR} = -\sum_{s=0}^{\infty}\text{diag}(Q_{i,t})^{-1}\frac{\partial Q_{i,t+s}}{\partial P_t'}\text{diag}(P_{i,t}) \tag{26}$$

Plugging (25) into (26) yields

$$\zeta_{i,t}^{LR} = \zeta_{i,t}^{SR}\left(1 + \sum_{s=1}^{\infty}\rho^s\right) = \frac{1}{1-\rho}\zeta_{i,t}^{SR} \tag{27}$$

I also define the short-run elasticity in absolute terms (instead of percentages) as $\tilde{\zeta}_{i,t}^{SR} = -\frac{\partial Q_{i,t}}{\partial P_{i,t}'}$.

[18] This implies that yesterday's holding in stock n affects both today's holding in n and m, but the spillover effect only comes through the short-term cross-elasticities, i.e. off-diagonal elements in $\zeta_{i,t}^{SR}$.

A.3 Equilibrium Price Impact

The market clearing condition implies that $\sum_{i=1}^{I} Q_{i,t} = Q^*$. We want to approximate the effects of an exogenous shock ΔX_t on current equilibrium prices at P_t and all future periods P_{t+s} for $s \to \infty$. I start with the instantaneous price impact. This is the same as the price impact in the static model with $\partial f_i / \partial Q'_{i,t-1} = 0$. Differentiating both sides of the market clearing condition with respect to X_t yields

$$\frac{\partial Q^*}{\partial X_t'} = \sum_{i=1}^{I} \frac{\partial f_i}{\partial X_t'} + \frac{\partial f_i}{\partial P_t'} \frac{\partial P_t}{\partial X_t'} \tag{28}$$

Shares outstanding are normalized to 1. Therefore $\frac{\partial Q^*}{\partial X'} = 0$. For simplicity, I define the elasticity in absolute terms (instead of percentages) as $\tilde{\zeta}_i^{SR} = -\frac{\partial f}{\partial P_t'}$. Rewriting (28) in terms of elasticities yields and solving for $\frac{\partial P_t}{\partial X_t'}$ yields

$$\frac{\partial P_t}{\partial X_t'} = \left(\sum_{i=1}^{I} \tilde{\zeta}_{i,t} \right)^{-1} \sum_{i=1}^{I} \frac{\partial f_i}{\partial X_t'}. \tag{29}$$

Let $\Delta d_t = \sum_{i=1}^{I} \frac{\partial f_i}{\partial X_t'} \Delta X_t$ be the first order approximation to the aggregate demand shock caused by the exogenous shock ΔX. Because shares outstanding are normalized to 1, Δd is measured in percent of total supply. A first-order approximation to the equilibrium increase in prices is given by $\Delta P = \left(\sum_{i=1}^{I} \tilde{\zeta}_i \right)^{-1} \Delta d$. Left-multiplying both sides by $\text{diag}(P_t)^{-1}$ and expressing the resulting return $\Delta p_t = \text{diag}(P_t)^{-1} \Delta P_t$ in terms of elasticities in percentages $\zeta_{i,t}$ yields the instantaneous (log) price change

$$\Delta p_t = \mathcal{M}_t^{SR} \Delta d_t, \tag{30}$$

where

$$\mathcal{M}_t^{SR} = \left(\sum_{i=1}^{I} \text{diag}(Q_{i,t}) \zeta_{i,t}^{SR} \right)^{-1}. \tag{31}$$

To simplify the interpretation of the long-run dynamics, I assume that the change in characteristics only affects the demand of an individual investor j with no inertia. Therefore, we can effectively treat the shock $\Delta d_t = \sum_{i=1}^{I} \frac{\partial f_i}{\partial X_t'}$ as a one-time change in supply without follow-on demand shocks due to inertia. The price impact from t until a future date $s > t$ is given by the ownership-weighted sum of elasticities corresponding to horizon

$h = s - t$. Formally, when inertia is given by $\frac{\partial f_i}{\partial Q'_{i,t+k-1}} = \rho_i \mathbf{I}$, this implies

$$\Delta p_{t:t+s} = \mathcal{M}_{t:t+s} \Delta d_t \tag{32}$$

where

$$\mathcal{M}_{t:t+s} = \left(\sum_{i=1}^{I} \text{diag}(Q_{i,t}) \zeta_{i,t} \frac{1 - \rho_i^{s+1}}{1 - \rho_i} \right)^{-1} \tag{33}$$

When $s = 0$, the dynamic multiplier is equal to the static short-run multiplier. When $s \to \infty$, the long-run multiplier $\mathcal{M}_t^{LR} \equiv \lim_{s \to \infty} \mathcal{M}_{t:t+s}$ is given by

$$\mathcal{M}_t^{LR} = \left(\sum_{i=1}^{I} \text{diag}(Q_{i,t}) \zeta_{i,t}^{LR} \right)^{-1}. \tag{34}$$

A.4 Counterfactual Experiments in KY (2019)

Motivated by the fact that portfolio weights are log-normally distributed in the data, KY (2019) use an exponential function for $f_i(P, X)$ and model investor demand as

$$w_i(n) = \frac{\exp(\delta_i(n))}{1 + \sum_n \exp(\delta_i(n))} \tag{35}$$

where $\delta_i(n) = \beta_{i,0} p(n) + u_i(n)$ and $u_i(n) = \sum_{k=1}^{K} \beta_{i,k} X_k(n) + \varepsilon_i(n)$. The sum of all investors must hold the market portfolio, thus

$$\log \sum_{i=1}^{I} A_i w_i(p) = p. \tag{36}$$

where $p \in \mathbb{R}^N$ is the vector of log market equities. Due to the exponential demand specification, equilibrium market equities are not available in closed form. However, a first-order approximation to the equilibrium price *change* following any counterfactual experiment is given by

$$\Delta p \approx \sum_{i=1}^{I} \left(\frac{\partial p}{\partial A_i} \Delta A_i + \frac{\partial p}{\partial u_i} \Delta u_i \right) \tag{37}$$

The demand shocks Δu_i capture demand changes due to changes in characteristics ΔX, preferences $\Delta \beta_i$ and latent demand $\Delta \varepsilon_i$. By differentiating the market clearing condition (36) with respect to the wealth distribution $\{A_i\}_{i=1}^I$ and demand shocks $\{u_i\}_{i=1}^I$ one can obtain closed-form solutions for the partial derivatives. I start with the assets under

management. The derivative of (36) with respect to the assets under management of investor j, A_j, is given by:

$$\frac{\partial p}{\partial A_j} = \mathbf{H}^{-1}\left(\sum_{i=1}^{I} A_i \frac{\partial w_i}{\partial p'}\frac{\partial p}{\partial A_j} + w_j\right), \tag{38}$$

where $\mathbf{H} = \sum_{i=1}^{I} A_i \text{diag}(w_i)$. Solving for $\frac{\partial p}{\partial A_j}$ yields

$$\frac{\partial p}{\partial A_j} = \underbrace{\left(\mathbf{I} - \sum_{i=1}^{I} A_i \mathbf{H}^{-1}\frac{\partial w_i}{\partial p'}\right)}_{\mathcal{M}} \mathbf{H}^{-1} w_j = \mathcal{M}\mathbf{H}^{-1} w_j \tag{39}$$

Similarly, the derivative with respect to the other demand shocks is given by

$$\frac{\partial p}{\partial u_j} = \mathcal{M}\mathbf{H}^{-1} A_j \frac{\partial w_j}{\partial d_j} \tag{40}$$

The partial derivatives of w_j with respect to p and d_j are available in closed form. Let $\mathbf{G}_j = \text{diag}(w_j)(\mathbf{I} - \mathbf{1}w_j')$. The partial derivatives can be expressed as $\frac{\partial w_j}{\partial p'} = \mathbf{G}_j \beta_{j,0}$ and $\frac{\partial w_j}{\partial d_j'} = \mathbf{G}_j$. Plugging into (37) yields

$$\Delta p \approx \mathcal{M}\left(\sum_{i=1}^{I} \mathbf{H}^{-1} w_i \Delta A_i + \mathbf{H}^{-1}\mathbf{G}_i \Delta u_i\right) \tag{41}$$

Note, that $\mathbf{H}^{-1}(w_i A_i)$ is the vector of ownership shares of investor i in all stocks. Because shares outstanding are normalized to 1, this is equivalent to $\text{diag}(Q_i)$. Thus we can express the equilibrium return as the product of a demand shock Δd and the multiplier matrix

$$\Delta p = \mathcal{M}\Delta d \tag{42}$$

where

$$\Delta d = \sum_{i=1}^{I} \text{diag}(Q_i)\left(\mathbf{1}\frac{\Delta A_i}{A_i} + (\mathbf{I} - \mathbf{1}w_i')\Delta u_i\right) \tag{43}$$

Because shares outstanding are normalized to 1, Δd is the aggregate demand shift in percent relative to total supply due to flows $\frac{\Delta A_i}{A_i}$ and portfolio changes $(\mathbf{I} - \mathbf{1}w_i')\Delta d_i$. Note, lastly, that as in KY (2019) the relationship between an investor's elasticity ζ_i and

β_i is given by

$$\zeta_i = \mathbf{I} - \beta_i(\mathbf{I} - \mathbf{1}w_i') \tag{44}$$

See equation (14) in their paper. Plugging into the expression for $\mathcal{M}$ in (39) yields

$$\mathcal{M} = \left(\sum_{i=1}^{I} \text{diag}(Q_i)\zeta_i\right)^{-1} \tag{45}$$

A.5 Microfoundation: Partial Adjustment Model

Consider investor i with risk aversion γ_i who maximizes a mean-variance objective by choosing a vector of portfolio weights $w_{i,t} \in \mathbb{R}^{\mathrm{N}}$. Subjective expected returns and covariances are given by $\mu_{i,t}$ and $\Sigma_{i,t}$ respectively. For simplicity, there are two periods $t-1$ and t. This is equivalent to assuming that the investor is myopic and maximizes one-period ahead risk-adjusted portfolio returns subject to transaction costs. The non-myopic extension can be found in e.g. Gârleanu and Pedersen (2013) or Bacchetta et al. (2023) but does not add incremental insights for this microfoundation. The investor faces portfolio adjustment costs ($AC_{i,t}$) when trading $\Delta w^a_{i,t} = w_{i,t} - w^p_{i,t}$ dollars, where $w^p_{i,t}$ are the passive buy-and-hold weights at time t in the absence of trading between $t-1$ and t. When assets are normalized to 1, $\Delta w^a_{i,t}$ are investor i's trades denoted in dollar units. Following Gârleanu and Pedersen (2013), I assume quadratic portfolio adjustment costs that are proportional to the covariance matrix of returns

$$AC_{i,t}(w_{i,t}) = \frac{c_{i,t}}{2}\left(\Delta w^a_{i,t}\right)' \Sigma_{i,t} \Delta w^a_{i,t}. \tag{46}$$

Quadratic adjustment costs are a direct consequence of transacting with linear market impact.[19] The investor maximizes the standard mean-variance objective

$$\max_{w_{i,t}} \left\{ w_{i,t}'\mu_{i,t} - \frac{\gamma_i}{2} w_{i,t}'\Sigma_{i,t} w_{i,t} - AC_{i,t}(w_{i,t}) \right\} \tag{47}$$

The optimal portfolio of shares at time t is a weighted mix of the passive buy-and-hold portfolio $w_{i,t}$ and the mean-variance optimal portfolio in the absence of trading costs

[19] Buying $\Delta w^a_{i,t}$ dollars worth of shares moves prices by $\Lambda_t \Delta w^a_{i,t}$ where Λ_t is a multidimensional version of Kyle's lambda. Trading costs are given by $\Delta w^a_{i,t}$ times the price impact resulting in the quadratic term $\left(\Delta w^a_{i,t}\right) \Lambda_t \Delta w^a_{i,t}$. If price impact is proportional to risk as in Greenwood (2005), this gives $\left(\Delta w^a_{i,t}\right) \Sigma_{i,t} \Delta w^a_{i,t}$.

$w^*_{i,t} = \frac{1}{\gamma_i}\Sigma^{-1}_{i,t}\mu_{i,t}$,

$$w_{i,t} = \rho_{i,t} w^p_{i,t} + (1-\rho_{i,t}) w^*_{i,t}, \tag{48}$$

where $\rho_{i,t} = \frac{c_{i,t}}{\gamma_i + c_{i,t}}$ captures inertia in the investor's portfolios. The current portfolio weight in stock n is an arithmetic mean of $w^p_{i,t}$ and $w^*_{i,t}$. However, when $w_{i,t}$ is not too far from $w^p_{i,t}$ and $w^*_{i,t}$, the arithmetic mean is close to the geometric mean $\left(w^p_{i,t}\right)^{\rho}\left(w^*_{i,t}\right)^{1-\rho}$. Pre-multiplying (48) by $\text{diag}(w_{i,t})^{-1}$ and writing at the stock-level gives

$$1 = \rho\frac{w^p_{i,t}(n)}{w_{i,t}(n)} + (1-\rho)\frac{w^*_{i,t}(n)}{w_{i,t}(n)}, \tag{49}$$

Subtracting 1 from both sides, and noting that $\log\frac{x}{y} \approx \frac{x}{y} - 1$ for small differences between x and y, we have

$$\log w_{i,t} = \rho \log w^p_{i,t} + (1-\rho)\log w^*_{i,t}. \tag{50}$$

Following Koijen and Yogo (2019), I assume that there is a factor structure in the covariance matrix of excess returns given by $\Sigma_{i,t} = \Gamma_{i,t}\Gamma'_{i,t} + \gamma_{i,t}\mathbf{I}$ and that factor loadings $\Gamma_{i,t}$ and expected returns $\mu_{i,t}$ are polynomial functions of stock-specific characteristics. Formally,

$$\Gamma_{i,t}(n) = y_{i,t}(n)'\mathbf{\Phi}_{i,t} + \phi_{i,t} \tag{51}$$

$$\mu_{i,t}(n) = y_{i,t}(n)'\mathbf{\Psi}_{i,t} + \psi_{i,t} \tag{52}$$

where $y_{i,t} = \begin{pmatrix} x_{i,t} \\ \frac{1}{2}\text{vec}(x_{i,t}x'_{i,t}) \\ \cdots \end{pmatrix} \in \mathbb{R}^{M\times N}$ is an Mth order polynomial and $x_{i,t} = [p_t, u_{i,t}]$ consists of log market equity p_t and other drivers $u_{i,t}$ capturing both observable and unobservable characteristics. The Woodbury matrix identity implies that

$$w^*_{i,t} = \frac{1}{\gamma}\Sigma^{-1}_{i,t}\mu_{i,t} = y_{i,t}\Pi_{i,t} + \pi_{i,t} \tag{53}$$

where $\pi_{i,t} \in \mathbb{R}^K$ does not vary across stocks. Under the coefficient restriction, that $\frac{\Pi_{i,t}}{w_{i,t}(0)} = \begin{pmatrix} \hat{\beta}_{i,t} \\ \frac{1}{2}\text{vec}(\hat{\beta}_{i,t}\hat{\beta}'_{i,t}) \\ \cdots \end{pmatrix}$ where $\hat{\beta}'_{i,t} = [\beta_{i,t}, 1]$, the target portfolio becomes exponential linear in log market equity p_t and $\log u_{i,t}$,

$$\delta^*_{i,t}(n) \equiv \log\frac{w^*_{i,t}(n)}{w_{i,t}(0)} = \beta_{i,t}\log p_t(n) + d_{i,t}(n), \tag{54}$$

where the demand shift $d_{i,t}(n) = \log u_{i,t}(n)$ captures demand for observable fundamentals as well as latent demand. Subtracting $\log w_{i,t}(0)$ from both sides of (50), yields

$$\delta_{i,t} = \rho_{i,t}\delta^p_{i,t} + (1-\rho_{i,t})\delta^*_{i,t}. \tag{55}$$

A.6 Short versus Long-Run Elasticities in Partial Adjustment Model

Plugging the long-run desired portfolio $\delta^*_{i,t}(n) = \beta_{i,t}p_{i,t}(n) + d_{i,t}(n)$ in (9) yields the following stock-level expression

$$\delta_{i,t}(n) = \rho_{i,t}\delta^p_{i,t}(n) + (1-\rho_{i,t})\beta_{i,t}p_{i,t}(n) + (1-\rho_{i,t})d_{i,t}(n). \tag{56}$$

Let $h_{i,t} = \log(A_{i,t-1}w_{i,t-1}(0))$ denote the log holdings in the outside asset. Note, that $\delta^p_{i,t}(n) = \log(Q_{i,t-1}(n)P_t(n)) - h_{i,t-1}$. For simplicity of notation, I assume no valuation gains in the outside asset. Any portfolio-level growth is subsumed by time-fixed effects in the empirical specification. Subtracting log prices and adding lagged holdings in the outside assets on both sides, we can rewrite the optimal portfolio as

$$q_{i,t}(n) = (1-\rho_{i,t})h_{i,t-1} + \rho_{i,t}q_{i,t-1}(n) + (1-\rho_{i,t})(\beta_{i,t}-1)p_{i,t}(n) + (1-\rho_{i,t})d_{i,t}(n) \tag{57}$$

Solving forward recursively results in the following long-run demand

$$q_{i,\infty}(n) = \frac{(1-\rho_{i,t})h_{i,t-1}}{(1-\rho_{i,t})} + \frac{(1-\rho_{i,t})(\beta_{i,t}-1)}{1-\rho_{i,t}}p_{i,t}(n) + \frac{(1-\rho_{i,t})d_{i,t}(n)}{1-\rho_{i,t}}. \tag{58}$$

Adding log prices $p_{i,t}$ on both sides and simplifying yields

$$\delta_{i,\infty}(n) = \beta_{i,t}p_{i,t}(n) + d_{i,t}. \tag{59}$$

Thus, over an infinitely long time horizon the investor holds the long-run desired portfolio $\delta_{i,\infty}(n) = \delta^*_{i,t}(n)$ with a long-run price response of $\beta_{i,t}$. Note, that $(1-\beta_{i,t})(1-\rho_{i,t})$ and $1-\beta_{i,t}$ are only approximations for the true short and long-run elasticities. The true elasticities are obtained by imposing the budget constraints $\sum_n w_{i,t}(n) = 1 - w_{i,t}(0)$ and $\sum_n w_{i,\infty}(n) = 1 - w_{i,\infty}(0)$ respectively, and differentiating $\delta_{i,t}(n)$ and $\delta_{i,\infty}(n)$ with respect

to current prices. This yields

$$\zeta_{i,t}^{LR} = \mathbf{I} - \beta_{i,t}\big(\mathbf{I} - \mathbf{1}w_{i,t}'\big) \tag{60}$$

$$\zeta_{i,t}^{SR} = \mathbf{I} - (\beta_{i,t}(1-\rho_{i,t}) + \rho_{i,t})\big(\mathbf{I} - \mathbf{1}w_{i,t}'\big) \tag{61}$$

where the long-run elasticity $\zeta_{i,t}^{LR}$ is the same as in KY (equation 14, 2019) and the short-run elasticity $\zeta_{i,t}^{SR}$ is roughly smaller by a factor $(1-\rho_{i,t})$. When investors hold diversified portfolios (i.e. the portfolio weights in individual stocks are small), the elasticity matrices simplify to $\zeta_{i,t}^{LR} \approx (1-\beta_{i,t})\mathbf{I}$ and $\zeta_{i,t}^{SR} \approx (1-\beta_{i,t})(1-\rho_{i,t})\mathbf{I}$ respectively.

Appendix B Estimation Details

B.1 Parameterizing Cross-Elasticities

Log-linearizing the general demand curve (1) yields the following first-order approximation

$$\Delta^h q_i(n) = -\sum_{m=1}^{N} \zeta_{i,nm}^{(h)} \Delta^h p(m) + d_i(n) \tag{62}$$

which is evaluating the percentage trade $\Delta^h q_i(n)$ as a response to a percentage change in the price $\Delta^h p(n)$ of all other stocks over different horizons h. I am omitting time t subscripts to avoid notational clutter. For trades obtained from quarterly 13F filings, the number of observations for each investor $(N \times T)$ is too small to estimate each element of the $N \times N$ elasticities matrix individually. However, the issue can be addressed by parameterizing $\zeta_{i,nm}^{(h)}$ as a function of observable stock-specific characteristics. The problem then reduces to estimating interaction effects as in Kelly et al. (2019). Logit demand (see KY, 2019) produce homogenous substitution patterns across stocks. Formally, logit substitution implies

$$\zeta_{i,nm}^{(h)} = \begin{cases} \zeta_i^{(h)}(1-w_i(n)), & m = n \\ -\zeta_i^{(h)} w_i(m), & m \neq n. \end{cases} \tag{63}$$

Cross-price elasticities are then subsumed by an investor-specific constant and (62) becomes a simple cross-sectional regression for each investor i given by (6) where $c_i^h = -\zeta_i^{(h)} \sum_{n=1}^{N} w_i(n)\Delta^h p(n)$ is the scaled portfolio return over h periods. Technically, $\zeta_i^{(h)}$ measures the difference between own- and cross-price elasticity. However, for diversified portfolios, this difference is negligible as individual portfolio weights are small.

B.2 Flow-Induced Trading

To construct exogenous price variation at a quarterly frequency, I use flow-induced trades by domestic equity mutual funds as in Lou (2012).[20]. I set quarterly flows less than -100% or greater than 200% to missing and only include funds for which the total assets computed from their portfolio holdings is between 75% and 120% of the total net assets (TNA) reported by CRSP. To the extent that fund flows are correlated with the funds' respective portfolio tilts, flow-induced trades may contain fundamental information and violate the exclusion restriction. For example, deteriorating earnings of technology firms may cause flows out of mutual funds with a technology tilt. I therefore construct orthogonalized flows $f_{i,t}^{\perp}$ using quarterly cross-sectional regressions of flows $f_{i,t}$ onto fund-level characteristics $C_{i,t}$. Formally,

$$\forall t : f_{i,t} = c_t + C_{i,t}\beta_t + f_{i,t}^{\perp} \tag{64}$$

where $C_{i,t}$ is a vector of fund-level characteristics. Fund-level characteristics are constructed using fund-level characteristic scores as in Lettau et al. (2018), which are portfolio-weighted averages of the stock characteristics $C_t(n)$ within a fund's portfolio, $C_{i,t} = \sum_{n=1} w_{i,t}(n)C_t(n)$. For every fund, I compute scores for greenness (based on MSCI ESG scores), value, size, momentum, profitability, investment, and beta.[21] Exogenous flow-induced demand is then simply given by summing over all hypothetical trades:

$$\Delta z_{i,t}(n) = \sum_{j \neq i}^{I} f_{j,t}^{\perp} Q_{j,t-1}(n) \tag{65}$$

Because the demand estimation is at the individual institution level, the correct instrument excludes all flow-induced trades from mutual funds that belong to institution i. Empirically, however, this does not materially affect any of the estimates because total flow-driven trading is not strongly affected by individual funds (even aggregated at the manager level). I therefore opt for the simpler version of the instrument given by $\Delta z_t(n) = \sum_{i=1}^{I} f_{i,t}^{\perp} Q_{i,t-1}(n)$. The long-horizon version of the flow-based instrument, $\Delta^h z_t$,

[20]The reduced-form literature on price multipliers has proposed several alternative exogenous demand shifts that could be used to instrument for price changes, such as index reconstitutions (see e.g. Chang et al. (2015), Pavlova and Sikorskaya (2022)) and, more recently, signed order flow aggregated to lower frequencies (Li and Lin (2022))

[21]Including dummy variables for all CRSP objective codes at the most granular level and controlling for past flow lags leaves all results in the paper unchanged.

is constructed by summing Δz_t over time.

B.3 Two-Stage Least Squares Estimation Over Different Horizons

I estimate investor-specific elasticities over different horizons using the flow-based instrument from 2000 to 2024 via two-stage least squares. I remove extensive trades (i.e. inclusions/exclusions from portfolios) and consider horizons h from 1 quarter up to 4 years. Due to limited observations from quarterly 13F filings and narrow holdings, elasticities are estimated over the panel of stock-level trades for investors with over 1,000 trades and over 3 years of data. Others, or those with weak first stages, are pooled into equal-sized bins by size. Formally, the estimator is given by

$$\begin{aligned} \text{1st Stage:} \quad & \Delta^h p_t(n) = m^{(h)} \Delta^h z_t(n) + u_t(n) \\ \text{2nd Stage:} \quad & \Delta^h q_{i,t}(n) = -\zeta_i^{(h)} \Delta^h \hat{p}_t(n) + d_{i,t}(n) \end{aligned}$$

Following KY (2019), I impose $\zeta_i^{(h)} > 0$ for all investors at all horizons. Note, that a decreasing first-stage price impact $m^{(h)}$ is not sufficient to argue for higher average elasticities at longer horizons. The key issue is that the price impact of a demand shift may decrease over time either because of increasing liquidity provision by arbitrageurs, i.e. increasing elasticities, or because the original demand shift was only temporary and slowly reverted. Explicitly estimating elasticities from investor-specific portfolio holdings allows disentangling these two effects. Beyond this, estimating investor-specific elasticities in a second stage has two distinct benefits. First, it gives insights into the underlying investor-specific determinants of the multiplier. Second, one can obtain stock-specific elasticities and time-varying effects even by estimating a scalar elasticity for every investor. Because ownership varies across stocks and time, ownership-weighted sums of elasticities vary across stocks and time.

B.4 First-Difference Estimator versus Logit Regressions

Motivated by the fact that investors' portfolio weights are log-normally distributed in the cross-section, Koijen and Yogo (2019) propose modeling portfolio weights as $w_i(n) = \frac{\exp(\delta_i(n))}{1+\sum_n \exp(\delta_i(n))}$ where

$$\delta_i(n) = \log \frac{w_i(n)}{w_i(0)} = \beta_i p(n) + u_i(n) \tag{66}$$

is estimated using quarterly cross-sectional regressions for each investor. Subtracting $p(n)$ from both sides and noting that the weight in the outside asset is subsumed by a constant c_i in a cross-sectional regression, this can be rewritten as

$$q_i(n) = -\zeta_i p(n) + c_i + d_i(n) \tag{67}$$

where $\zeta_i = 1 - \beta_i$.[22] The reduced-form estimator (6) for horizon h is hence a straightforward first-difference version of the level specification in KY (2019), with differences computed over h periods. The elasticities obtained from the two estimators may diverge for two concrete reasons. First, trades introduce a specific time-series dimension that is not present in the static logit estimation. Second, trades effectively control for lagged portfolios and hence remove hidden investment mandates that may lead to biased elasticities (if correlated with market equity). Appendix Section C illustrates the two effects in a simulated laboratory environment. The structural model presented in Section (6) allows disentangling the two effects empirically.

B.5 Instruments for lagged portfolios

Under persistent demand shocks $q_{i,t-1}(n)$ is potentially endogenous, i.e., $\text{cov}(\varepsilon_{i,t}(n), q_{i,t-1}(n)) \neq 0$. It is possible that the normalization by shares outstanding does not fully remove the effect of persistent demand shocks on both the current portfolio and lagged ownership. I, therefore, propose two simple instruments denoted by $z^q_{i,t-1}(n)$.

Long-run lagged Ownership.
First, I instrument $q_{i,t-1}$ by its 3-year lag. This instrument satisfies the exclusion restriction if persistent signals only affect demand for a limited amount of time (here <3 years). While this instrument is clearly imperfect, the exogeneity can be gauged by how much the resulting inertia coefficient changes. Moving from a quarterly to a 3-year lag does not materially affect the coefficients suggesting that long-run persistent demand shocks do not pose a strong potential bias.

Model-based Instrument.
Second, I propose a simple model-based instrument. Let $\delta^X_{i,t-1}(n)$ denote the fitted

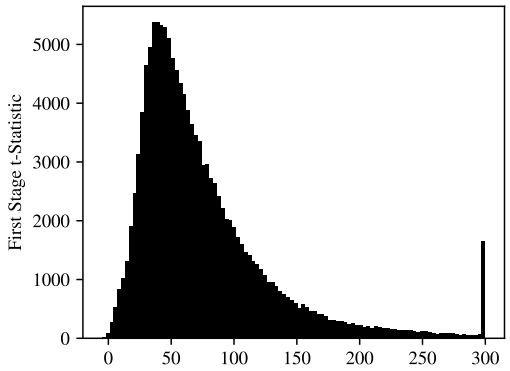
[22]The budget constraint, $\sum_n w_i(n) = 1 - w_i(0)$, implies own-price elasticities are given by $-\frac{\partial q_i(n)}{\partial p(n)} = \zeta_i(1 - w_i(n))$ as opposed to ζ_i. However, since $w_i(n)$ is small empirically, this adjustment has a negligible impact on the implied elasticities.

value from a cross-sectional regression of portfolios $\delta_{i,t-1}(n)$ onto characteristics *excluding* lagged ownership $q_{i,t-2}(n)$.[23] I then construct the instrument for $q_{i,t-1}(n)$ as the counterfactual ownership (dollar holdings relative to market cap) in the absence of latent demand shifts, formally $z^q_{i,t-1}(n) = \delta^X_{i,t-1}(n) - p_{t-1}(n)$. Note, this is similar to instrumenting $q_{i,t-1}(n)$ directly with all lagged characteristics $X_{t-1}(n)$ and the lagged mandate-based instrument $z^M_{i,t-1}(n)$. The exclusion restriction holds under the assumption that latent demand is not driven by lagged characteristics, i.e. lagged characteristics only affect current demand via the inertia channel. A violation of this assumption could be momentum demand to the extent that momentum is not subsumed by profitability (see Hou et al. (2015)). Controlling for momentum does not materially affect the resulting inertia coefficient.[24]

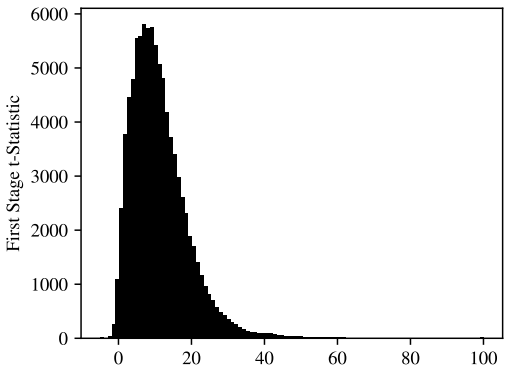
Figure B.1 plots the first-stage t-statistic for both instruments over all investor-quarter pairs. Both instruments are highly relevant for the vast majority of investor-quarter pairs. Pooling investor-quarter pairs with weak instruments and re-estimating their demand leaves all results unchanged.

Figure B.1: **Relevance Inertia Instrument**.
The figure plots the first stage t-statistic on the model-based instrument (Panel a) and long-run lagged ownership (Panel b) for all investor-quarter pairs. The first-stage regression is given by $q_{i,t-1}(n) = \beta_{i,t} z^q_{i,t-1}(n) + \mathcal{M}_{i,t} z^q_{i,t}(n) + Controls + \epsilon_{i,t}(n)$ where the controls include log book equity, dividends to book equity, profitability, investment, and market beta as of t, as well as an investor-time-specific constant.

(a) Model-based Instrument



(b) Long-run lagged Ownership



[23] Effectively, $\delta^X_{i,t-1}(n)$ is the fitted value from the level specification in KY (2019).
[24] The results are available upon request.

B.6 Constrained Estimation of Partial Adjustment Model

Putting the instruments and coefficient constraints together, the moment condition for identifying the partial adjustment model is given by

$$\begin{aligned} &\mathbb{E}_t\big[\varepsilon_{i,t}(n) \mid z^q_{i,t-1}(n), z^M_{i,t}(n), X_t(n)\big] \\ &\qquad \text{s.t.} \quad \rho_{i,t} \in [0,1), \alpha_{i,t} < 1. \end{aligned} \tag{68}$$

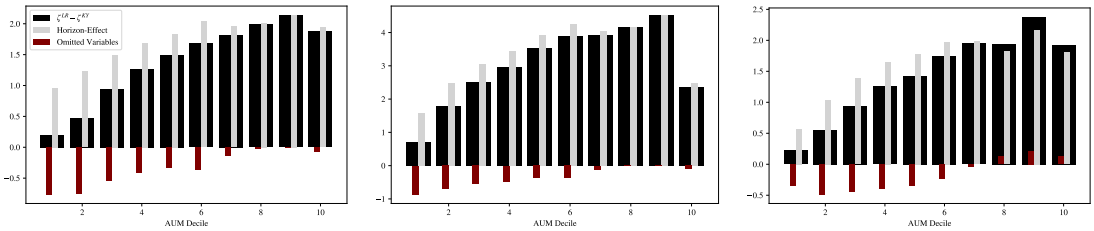
where $\varepsilon_{i,t}(n) = \delta_{i,t}(n) - \rho_{i,t} q_{i,t-1}(n) - \alpha_{i,t} p_t(n) - \sum_{k=1}^{K} X_{k,t}(n)\beta_{i,t,k}$. I estimate (68) via constrained GMM. However, estimation via unconstrained OLS and imposing the coefficient constraints ex-post yields roughly equivalent results.

B.7 Results for different instruments

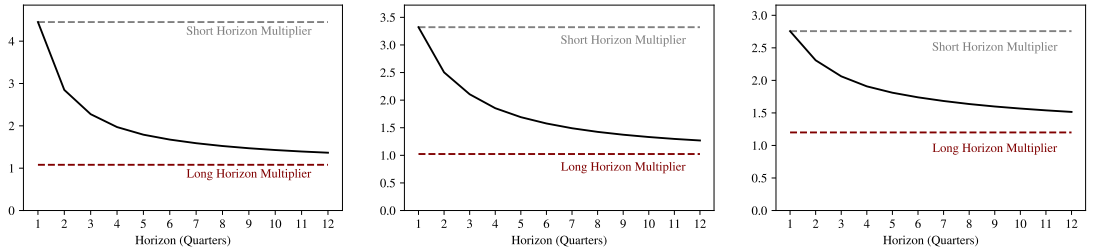
Figure B.2 illustrates the robustness of the results with respect to different instruments. Regardless of the instrument, investors are inert with an average $\rho_{i,t}$ of around 0.65-0.85 (see Table 2). Regardless of the instrument, the coefficient difference between holding and trading $\zeta^{SR}_{i,t} - \zeta^{KY}_{i,t}$ is positively correlated with investor size (Panel b). Short-run multipliers greatly exceed long-run multipliers and are of the same magnitude across instruments (Panel c). The reversal implied by the long-run elasticities is in line with the impact reversal from flow-induced trades for all instruments (Panel d). Last, conditioning on model-based reversal $\mathcal{R}_t(n)$ improves unconditional momentum returns for all instruments.

Figure B.2: **Robustness**.
The figure reports all results for different instruments for $q_{i,t-1}(n)$. The left-most column reports the baseline OLS results without instruments for $q_{i,t-1}(n)$. The middle column reports results for the model-based instrument. The right-most column reports results for using long-run lagged ownership as an instrument. Panel a) plots the distribution of the coefficient difference between holding and trading $\zeta_{i,t}^{SR} - \zeta_{i,t}^{KY}$. The remaining panels plot all results from Section 7 and 8 for the different instruments: Panel b) plots the median short and long-run multipliers, Panel c) reports the out-of-sample test, and Panel d) plots momentum returns conditional on model-based reversal $\mathcal{R}_t(n)$.

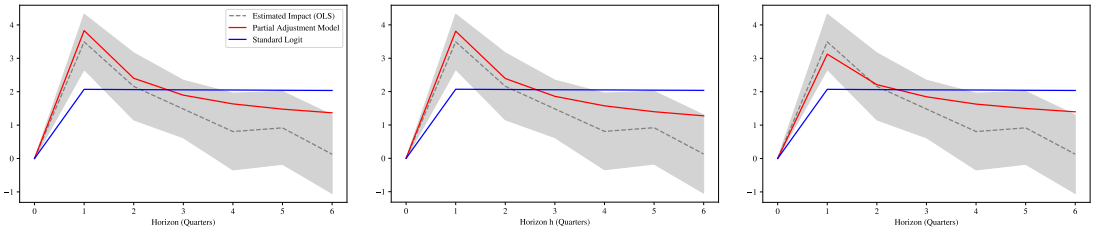
(a) Distribution $\zeta_{i,t}^{SR} - \zeta_{i,t}^{KY}$



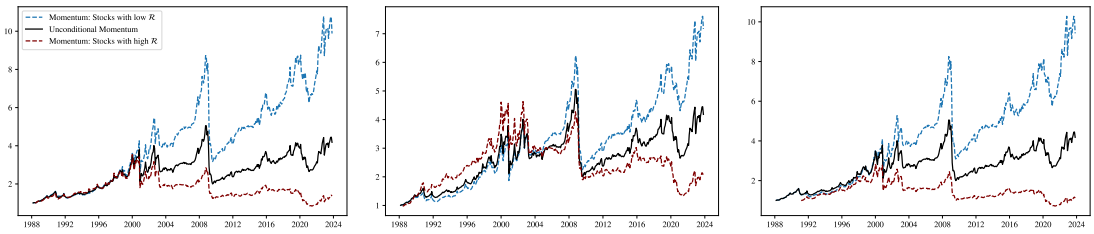
(b) Short versus Long-run Impact



(c) Predicting the Impact of Flow-Induced Trades



(d) Inertia $\mathcal{R}_t(n)$ and Momentum Returns



Appendix C Simulation

To illustrate the effects of the omitted variable bias (b) and trading horizon (h), I provide simple simulations from two stylized investors $i = \{b, h\}$. Both investors enter the market in $t = 2010$ with \$1 under management. Investor b (for bias) is used to illustrate the potential omitted variable bias from unobservable portfolio tilts in the cross-section. Investor h (for horizon) is used to illustrate the effects of trading horizons on the estimated elasticities. I simulate their portfolios using observable market data from 2010 to 2024 and estimate their elasticities cross-sectionally in reduced form (i.e. from changes) and structurally (i.e. from levels). Formally, the regression specifications are given by $\log w_{i,t}(n) = \beta_{i,t} p_t(n) + c_t + \epsilon_{i,t}(n)$ and $\Delta \log w_{i,t}(n) = \beta_{i,t} \Delta p_t(n) + c_t + \epsilon_{i,t}(n)$. To this end, I construct a balanced panel of stock-quarter observations by taking the subset of all US stocks as of January 2024 that were already publicly traded in 2010. At every quarter t, the cross-section of both investors' (log) portfolio weights can be decomposed linearly

$$\log w_{i,t}(n) = \beta_i p_t(n) + \epsilon_{i,t}(n) \quad \text{for} \quad t \geq 2010. \tag{69}$$

where $\epsilon_{i,t}(n)$ are unobserved demand shifts coming from e.g. differential starting portfolios or dependence on lagged weights (inertia). β_i measures the slope of the demand curve, i.e. the cross-sectional relationship between portfolio weights and prices. The OLS estimate of $1 - \beta_i$ is an unbiased estimator of the elasticity of investor i as long as $\epsilon_{i,t}(n)$ and $p_t(n)$ are cross-sectionally uncorrelated.

C.1 Investor b: Omitted variables Bias

Investor b enters the market with portfolio tilted towards technology stocks. For simplicity, assume that as of $t = 2010$ the technology tilt perfectly coincides with market equity. Thus she enters the market with a value-weighted portfolio. Formally, her starting portfolio is given by $\log w_{b,2010}(n) = p_{2010}(n) - p_{2010}$ where $p_{2010} = \log \sum_{n=1}^{N} P_{2010}(n)$ is the log total market cap as of 2010. She then has no other demand shifters, and henceforth trades *quarterly* against price moves with an elasticity of 1, thereby keeping her portfolio weights constant at their starting level $w_{b,2010}(n)$. In the generic demand curve (69), investor b's starting portfolio implies that $\epsilon_{b,t}(n) = p_t(n) - p_{2010}$ and $\beta_{b,t}(n) = 0$. Estimating elasticities from portfolio holdings in levels leads to a coefficient bias $\zeta_{b,t}^{\text{Levels}} - \zeta_{b,t}$ that is driven by how cross-sectionally correlated current market equities are with the starting

values $p_{2010}(n)$ in 2010. Formally, the coefficient bias is given by

$$\zeta_{b,t}^{\text{Levels}} - \zeta_{b,t} = \frac{\text{cov}_t(p_{2010}(n), p_t(n))}{\text{var}_t(p_t(n))}, \tag{70}$$

which tends to be positive because cross-sections of market equities are quite persistent. When $t = 2010$ the coefficient bias is exactly 1. In first differences however, $\Delta\epsilon_{b,t}(n) = 0$, and $\zeta_{b,t}^{\text{Changes}} - \zeta_{b,t} = 0$. Figure C.3 illustrates the relative advantages of the estimation in levels and changes. Investor b's elasticity in levels is "perfectly" biased. She starts with a value-weighted portfolio that is *cross-sectionally* passive but trades *actively* against price movements. Over time, the initially value-weighted portfolio moves further and further away from true market prices resulting in a smaller coefficient bias. However, cross-sections of market equities are so persistent, that a substantial bias remains even after 12 years. In first differences, the bias from the omitted variable (the value-weighted starting portfolio) is exactly removed.

C.2 Investor h: Trading Horizon

Investor h enters the market with an equal-weighted portfolio. Hence, there is no omitted variable bias. However, instead of assuming that she trades quarterly against price movements, she slowly trades back to her target portfolio at a rate $1 - \rho$. Formally, her starting portfolio is given by $\log w_{h,2010}(n) = \log(1/N)$. Because she slowly trades back to her target portfolio at a rate $1 - \rho$, her portfolio weights are given by $\log w_{h,t}(n) = \rho \log w_{h,t}^p(n) + (1 - \rho)\log(1/N)$, where $w_{h,t}^p(n) = w_{h,t-1}(n)\frac{P_t(n)}{P_{t-1}(n)}$ is the passive buy-and-hold portfolio the investor would hold in the absence of trading. Solving this recursive equation yields $\log w_{h,t}(n) = \sum_{s=1}^{t} \rho^{t-s+1} \log \frac{P_s(n)}{P_{s-1}(n)} - N$. For simplicity, consider the simple case of a single price change directly after $t = 2010$ and no price changes thereafter. The portfolio weight at any time $t \geq 2010$ is given by $\log w_{h,t}(n) = \rho^{(t-2010)}(p_t(n) - p_{2010}(n)) - \log N$ and the coefficient bias from portfolio holdings in levels is

$$\zeta_{h,t}^{\text{Levels}} - \zeta_{h,t} = \rho^{(t-2010)} \left(1 - \frac{\text{cov}_t(p_{2010}(n), p_t(n))}{\text{var}_t(p_t(n))}\right). \tag{71}$$

As long as prices remain close to their starting values, i.e. $\frac{\text{cov}_t(p_{2010}(n), p_t(n))}{\text{var}_t(p_t(n))} \approx 1$, the level specification gives an unbiased estimate of the true long-run elasticity $\zeta_{h,t} = 1$. Similarly, if prices change but $\rho \to 1$, the level specification remains unbiased. Estimating

elasticities from trades, however, greatly understates the true long-run elasticity. To this end, note that changes in holdings are given by

$$\Delta \log w_{h,t}(n) = \rho \Delta p_t(n) + (1-\rho) \sum_{s=1}^{t-1} \rho^{t-s} \Delta p_s(n) \tag{72}$$

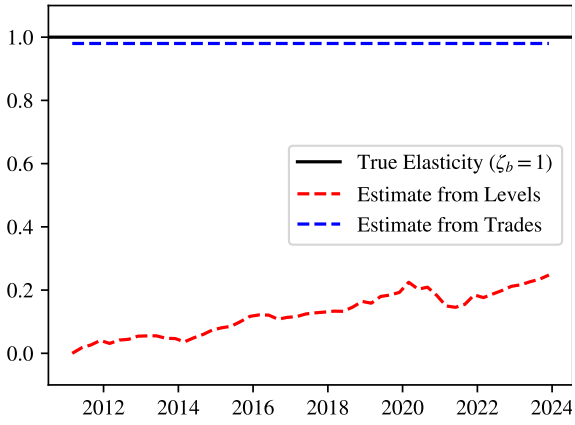
For simplicity, assume that there is no cross-sectional correlation in returns. Cross-sectional regressions of $\Delta \log w_{h,t}(n)$ onto $\Delta p_t(n)$ therefore yield an implied elasticity of $\zeta_{h,t}^{\text{Changes}} = 1 - \rho$. The coefficient bias is given by

$$\zeta_{h,t}^{\text{Changes}} - \zeta_{h,t} = \rho \tag{73}$$

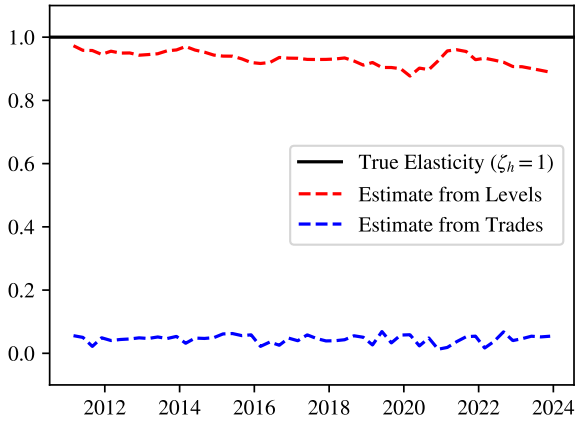
The slower the investor trades, the larger the coefficient bias. In the limit, as $\rho \to 1$, the investor trades infinitely slow and the estimation in changes is "perfectly" biased. The right panel of Figure C.3 shows that the level specification captures the true long-term elasticity of 1. The specification in changes, however, only captures the quarterly short-term responses of $1 - \rho$. Therefore, while investor b's elasticity is better captured using trades, investor h is better captured using levels.

Figure C.3: **Omitted Variable Bias versus Trading Horizon**.
The figure plots the estimated elasticities from simulated portfolio data using true market prices from 2010 to 2024. I simulate the portfolio holdings $q_{i,t}(n)$ for two investors $i = b, h$ with equal (long-run) elasticities $\zeta_b = \zeta_h = 1$. I estimate their elasticities quarterly from cross-sectional regressions in levels, i.e. $q_t^s(n) = -\zeta_t^s p_t(n) + c_t + \epsilon_t(n)$, and quarterly cross-sectional regressions in changes, i.e. $\Delta q_t^s(n) = -\zeta_t^s \Delta p_t(n) + c_t + \epsilon_t(n)$. Panel a) plots the estimated elasticities for investor b, who enters the market in 2010 with a value-weighted portfolio and trades back to her starting portfolio quarterly. Panel b) plots the estimated elasticities for investor h, who enters the market in 2010 with an equal-weighted portfolio and gradually trades back to her starting portfolio with speed $\rho = 0.95$.

(a) Omitted Variable Bias



(b) Trading Horizon



Appendix D Additional Figures

Figure D.4: **Inertia and Passive-Portfolios: Robustness to Extensive Trades**
The figure plots R^2s from univariate cross-sectional regressions of investor-specific portfolio weights onto a passive buy-and-hold portfolios $w^p_{i,t}(n) = \frac{Q_{i,t-h}(n)P_t(n)}{\sum_n Q_{i,t-h}(n)P_t(n)}$ with h ranging from 1 to 8 quarters. I construct current and lagged portfolios over the subset of all stocks held with a positive weight in at least one of the two portfolios. Whenever a currently held stock is missing in the lagged portfolio, I set its passive portfolio weight equal to 0. When a stock is held in the lagged portfolio but currently missing, I set its current portfolio weight equal to 0. For each institution-quarter pair, I estimate $w(n) = \alpha + \beta w^p(n) + \epsilon(n)$ and report the distribution and average R^2 for each horizon.

