Erasing Alpha[*]

Carter Davis[†] Xiao Han[‡] Stanislav Sokolinski[§] Andrea Tamoni[¶]

February 2026

Abstract

How much capital must move to eliminate the alpha of an investment strategy? We address this question within a stock-level equilibrium framework that delivers a transparent decomposition of alpha erosion into three components: price impact, its pass-through into expected returns, and the ownership overlap between the targeted portfolio and the source and destination of capital. Focusing on the value strategy, we find that alpha is remarkably hard to erase: even reallocating *the entire* growth-fund AUM into value funds reduces value alpha from 5.6% to only 4.9%. This modest erosion implies a value capacity of about \$3.1 trillion, and alternative funding sources yield even larger capacity. The same reallocation generates a whack-a-mole spillover by *raising* momentum alpha. Our results suggest that even capital flows large enough to generate meaningful price impact translate into only modest changes in alphas, and that overlapping portfolio holdings can mechanically redistribute alpha across strategies rather than eliminate it.

Keywords: Factor Investing, Institutional Investors, Market Efficiency.
JEL codes: G11, G12, G23.

[*]An earlier version of this paper circulated under the title "Rewriting Expected Returns." We thank Vikas Agarwal, Amit Goyal, Hao Jiang, Yukun Liu, Ye Li, Marcel Müller, Anna Pavlova, Shri Santosh, Hanbin Yang, and seminar and conference participates at participants at the 2025 Conference on New Perspectives and New Products in Asset Management at Bayes Business School, the 2025 SAFE Asset Pricing Conference, and the 2025 CFEA Conference for valuable comments.

[†]The Ohio State University Fisher College of Business. Address: Department of Finance, 2100 Neil Ave, Columbus, OH 43210. E-mail: carter.davis@fisher.osu.edu. Web: https://sites.google.com/site/carterkentdavis/.

[‡]Bayes Business School, City University of London. Address: 106 Bunhill Row, London EC1Y 8TZ. E-mail: xiao.han.3@city.ac.uk. Web: https://sites.google.com/site/xiaohanfinance

[§]Broad College of Business, Michigan State University. Address: Department of Finance, 315 Eppley Center, East Lansing, MI 48824. E-mail: sokolins@msu.edu. Web: https://stanislavsokolinski.wordpress.com.

[¶]Mendoza College of Business, University of Notre Dame. Address: 304 Mendoza College of Business, Notre Dame, IN 46556. E-mail: andrea.tamoni.research@gmail.com. Web: https://www.andreatamoni.com.

1 Introduction

Modern equity markets are dominated by delegated and increasingly systematic capital. Mutual funds, ETFs, and quantitative investing shift trillions of dollars across broad styles and long–short strategies at a scale that would have been hard to imagine when many alpha-bearing strategies were first documented. This evolution turns a classic asset-pricing observation into a concrete equilibrium question: if abnormal returns are an opportunity, how much capital must move to erase them, and why do many alphas remain stubbornly persistent in the face of industrial-scale trading? A naïve arbitrage view predicts rapid compression as capital reallocates. In practice, alphas survive, implying that the effect of reallocations can be weak—and, when strategies overlap in holdings, can resemble a whack-a-mole problem in which compressing one alpha mechanically inflates another.

A useful way to see why is to recognize that capital reallocations do not "attack" a strategy directly. Instead, they reshuffle ownership across underlying stocks, shift marginal holders, and rewrite expected returns through market clearing. Alpha erosion is therefore inseparable from the distribution of ownership and from how investors substitute across stocks when relative prices adjust.

This perspective differs from prominent equilibrium theories of delegated management that define alpha at more aggregate levels. In Berk and Green (2004), alpha is a *fund-level* outcome competed away as assets under management grow. In Pástor and Stambaugh (2012), decreasing returns to scale operate at the *industry* level, diluting performance as the active sector expands. While these studies provide foundational discipline on how performance erodes with scale, they do not directly address how a given capital reallocation reshuffles stock-level demand and, in turn, rewrites portfolio alpha. That mapping is our goal: we quantify how reallocations reprice the underlying stocks, how much of that repricing passes through into alphas, and how the effects propagate to other strategies through overlapping holdings.

Our core object is a decomposition of alpha erosion into three economically interpretable components:

$$\Delta\alpha \approx \underbrace{\big(\text{price impact per \$}\big)}_{\text{stock-level multiplier}} \times \underbrace{\big(\text{pass-through}\big)}_{\text{stock-level price}\rightarrow\text{expected return}} \times \underbrace{\big(\text{ownership overlap}\big)}_{\text{reallocation exposure from holdings}} . \quad (1)$$

The first term is the stock's equilibrium *price impact per dollar*: how far its price must adjust to clear the market given downward-sloping demand. The second term is *pass-through*: the fraction of that price change that translates into a change in expected returns - persistent price impact does not compress long-term alpha, whereas transitory repricing does. Both objects are stock-level and must be estimated. The third term is *ownership overlap*, constructed from observed holdings: it maps a reallocation into the stock-level effects that aggregate into a strategy-level alpha change by combining who loses capital, who gains it, and how their portfolios differ. If the funding side does not own the stocks that need to be sold to compress returns, the induced selling is misaligned and alpha erodes slowly even when the destination tilts strongly toward the strategy.

We estimate the price-impact term from a market-clearing demand system by recovering stock-specific multipliers that map a dollar demand shock into an equilibrium price change (Koijen and Yogo, 2019; Koijen et al., 2023). Because price impact can depend on how investors substitute across assets, we allow flexible substitution by including cross-substitution among economically similar stocks (Fuchs et al., 2023; Chaudhary et al., 2025; Li and Lin, 2025). We validate this model-based price impact against realized flow-based shocks studied in the price-pressure literature (e.g., Coval and Stafford, 2007; Frazzini and Lamont, 2008; Lou, 2012; Wardlaw, 2020). Using mutual-fund trading shocks and the quasi-experimental 2002 Morningstar ratings change, model-implied price effects align closely with reduced-form return effects at both the stock and portfolio level.

We estimate pass-through using only the present-value accounting identity. A price change erodes alpha only to the extent it reflects discount-rate news rather than cash-flow news, and the present-value restriction pins down that mapping through the persistence of repricing. Using the classic Campbell–Shiller decomposition we estimate a single stock-level pass-through parameter that converts repricing into expected-return changes (Campbell and Shiller, 1988; Cochrane, 2008). This step does not require modeling demand dynamics and cleanly separates "how much prices move" from "how much expected returns move."

With these ingredients in hand, we turn to our main object: how capital reallocation erodes the alpha of the value strategy. We center on value because it is the canonical long-horizon trade with economically large alpha and because it maps directly into the style buckets that organize delegated capital. We then track momentum under the same reallocations to quantify *cross-capacity*: value and momentum load on overlapping sets of stocks, so reallocations that crowd into value mechanically reprice momentum winners and losers as an unintended byproduct (Jegadeesh and Titman, 1993; Asness, 1994).

We study two counterfactual reallocations that hold the destination fixed—value funds receive the same dollar inflow—and vary only the funding side, thereby isolating the source-of-capital channel in equilibrium. The first reallocation shifts capital from growth funds to value funds, capturing a classic style rotation (Barberis and Shleifer, 2003). Because growth funds concentrate in the growth leg of value while value funds concentrate in the value leg, this rotation translates mechanically into buy–sell pressure aligned with the value spread. The second sends the same dollars into value funds but funds them with household outflows, capturing the well-documented transition from direct investing to delegation (Stambaugh, 2014). Studying these reallocations lets us quantify how quickly value alpha erodes when value is crowded, whether the same capital shift generates spillovers to momentum through overlapping holdings, and how the answer changes with the source of capital.

Our headline result is that value alpha is extremely hard to erase, and attempts to do so can *raise* momentum alpha. In the baseline, the value strategy earns a CAPM alpha of 5.6% per year. Reallocating an amount equal to 100% of aggregate growth-fund AUM into value funds reduces value alpha by only 0.7%, leaving a post-reallocation alpha of 4.9%. The point is not that value is immune to crowding—prices move in the expected direction—but that even very large, economically salient reallocations translate into surprisingly modest alpha erosion in equilibrium.

That weakness becomes stark once we translate it into *capacity*, defined as the dollar reallocation required to drive value CAPM alpha to zero. Our baseline estimate is about $3.1 trillion when funded by growth outflows— nearly 900% of aggregate growth-fund AUM. This magnitude reconciles two ideas that are often discussed in isolation: reallocations can move prices in inelastic markets, yet abnormal expected returns can remain persistent because eliminating alpha requires reallocations that are simply too large, too coordinated, and too concentrated in the relevant stocks to occur in practice.

The same growth-to-value reallocation also produces a sharp spillover to momentum: momentum CAPM alpha rises by 0.4%, from 16.9% to 17.3%. This is *cross-capacity* in its cleanest form: because momentum winners overlap heavily with growth stocks (Jegadeesh and Titman, 1993; Asness, 1994), pulling capital out of growth shifts relative demand against winners more than losers, mechanically widening the momentum spread. The implication is a whack-a-mole effect: when strategies share the same underlying names, compressing one spread can mechanically inflate another, so arbitrage capital directed at one alpha can unintentionally create or strengthen a different one.

Finally, the source of capital is first order. Funding the same inflow to value funds with household outflows instead of growth-fund outflows produces much smaller effects: value CAPM alpha falls by only 0.4% (5.6% to 5.2%), and the momentum spillover is negligible (about 0.1%). The implied value capacity is therefore much larger, about $5.0 trillion, roughly 95% of household equity AUM. The lesson is that dollars are not fungible: what matters is not only where capital goes, but which portfolios are shrunk to fund it. Because households hold broad market exposure, their outflows are spread across many stocks, generating diffuse selling that is poorly aligned with the value portfolio. As a result, expected returns rewrite much more slowly, underscoring that investor identity has a first-order effect on how strategy returns evolve (Li et al., 2026).

Why is capacity so large? Equation (1) makes the constraints explicit. Capacity is high when aggregate demand is elastic, when source–destination holdings have limited overlap with the anomaly portfolio, and when pass-through is small. These forces interact: ownership can be concentrated enough for reallocations to generate meaningful price pressure, yet alphas move only with the persistent expected-return component of that repricing. The broader message is that the survival of alpha-bearing strategies in the era of systematic investing is an equilibrium question about ownership, price impact, and pass-through: reallocations can move prices and still leave expected-return wedges largely intact, and overlapping holdings can mechanically reassign alpha across strategies.

Importantly, our findings do not adjudicate between risk-based and mispricing-based interpretations of anomalies. In a risk-based view, reallocations compress premia by shifting who is willing to bear a given exposure at lower required returns; in a mispricing-based view, they expand arbitrage capital and correct misvaluation. We are deliberately agnostic about the motive behind alpha–preferences, beliefs, mandates, or frictions. What we pin down is the equilibrium *technology* of alpha erosion, not a particular theory of why alphas exist. In this sense, sufficiently large reallocations can eliminate alpha regardless of its root cause, and persistence is naturally traced to the scale, composition, and spillovers of reallocations rather than to any single explanation.

Our approach also differs from the classic capacity literature: we study *reallocation ca-*

pacity in equilibrium, whereas papers such as Korajczyk and Sadka (2004) study *implementation capacity* in trading. Their object is scalability under turnover and execution costs—how much capital can be deployed before market impact and transaction costs wipe out *net-of-costs* returns. Ours is equilibrium feasibility—how much capital must be shifted across investor groups to reprice the underlying stocks enough to compress the strategy's *gross* alpha. The distinction matters because a premium can persist for two different reasons: because it is too costly to harvest, or because eliminating it would require reallocations that are implausibly large or poorly aligned with who holds the relevant stocks.

Our work is related to the large literature on price pressure and demand shocks that uses realized reallocations—index reconstitutions, fire sales, and mutual fund trading—to measure how forced buying and selling moves prices (e.g., Coval and Stafford, 2007; Frazzini and Lamont, 2008; Lou, 2012; Chang et al., 2015). We draw on these environments for validation because they deliver ownership shocks with an empirically measurable price response, but our object is different: rather than estimating reduced-form price effects of realized events, we use an equilibrium mapping to quantify how large reallocations rewrite alphas, including spillovers across overlapping portfolios. In doing so we also speak to the demand-system and inelastic-markets agendas, which emphasize that reallocation can be first-order for prices even in deep markets (Koijen and Yogo, 2019; Koijen et al., 2023; Gabaix and Koijen, 2022). We make progress by translating that insight into a strategy-level equilibrium quantity—reallocation capacity—and to show how it depends on source-of-capital constraints and overlap-driven spillovers.

Recent work emphasizes that price impact is dynamic and horizon-dependent. Van der Beck (2025) shows that demand becomes substantially more elastic in the long run, so short-horizon price impact can partially unwind as investors rebalance, and Van der Beck et al. (2025) study how flow-induced price pressure can mechanically inflate realized fund returns and fuel return-chasing feedback. These papers sharpen the time-series anatomy of price pressure—impact, amplification in realized performance, and reversal. Our angle is different: we treat ownership shifts as equilibrium primitives and ask a comparative-static question about counterfactual alphas under explicit capital reallocations. In our framework, the dynamic considerations enter only through two transparent sufficient statistics—the relevant price-impact multiplier and the pass-through from repricing to expected returns—while the central focus is on alpha and reallocation capacity as equilibrium objects.

2 Theoretical Framework

This section develops a demand-based framework for answering a single question: when capital is reallocated across investor groups, how does the market re-price the cross-section in equilibrium, and what does that repricing do to an investment strategy's alpha? We proceed in three steps. The first two steps operate at the stock level, and the third step aggregates to the portfolio level.

First, Section 2.1 maps an investor-level capital reallocation into equilibrium price impact using a demand system with a full cross-asset structure, so reallocations can generate spillovers across assets. Second, Section 2.2 translates these price changes into expected returns using a Campbell–Shiller pass-through logic, which recognizes that a price change does not mechanically translate one-for-one into expected returns. Third, Section 2.3 aggregates the implied stock-level expected-return changes into portfolio-level effects, delivering transparent expressions for changes in portfolio alpha and a natural notion of portfolio *capacity*: the dollar reallocation required to drive a given alpha to zero. We keep the main text focused on the economic mapping and intuition; derivations and technical details are in Appendix B.

2.1 Price Impact

Investor demand Let i index investors and let $n \in \{1, \dots, N\}$ index assets. Let $P_{n,t}$ denote the per-share price and $p_{n,t} \equiv \log P_{n,t}$ the log price. Let $A_{i,t}$ denote investor i's AUM and $a_{i,t} \equiv \log A_{i,t}$. If investor i allocates portfolio weight $w_{n,i,t}$ to asset n, then its share demand is the corresponding dollar position divided by the price,

$$Q^d_{n,i,t} = \frac{w_{n,i,t} \, A_{i,t}}{P_{n,t}}. \tag{2}$$

We model how these holdings respond to prices using a generalized isoelastic demand system with a full cross-elasticity matrix. Isoelastic demand is our maintained empirical specification for two reasons. First, it provides a tight approximation to the Koijen–Yogo logit mapping used in prior work (Davis, 2024; Koijen and Yogo, 2019). Second, it can be derived as a disciplined benchmark from classical preferences and delivers empirically reasonable price-impact magnitudes for capital reallocation across investors and assets (Davis, 2025).

The key object is investor i's elasticity matrix $\mathbf{E}_{i,t} \in \mathbb{R}^{N\times N}$, defined entry-by-entry as

$$(\mathbf{E}_{i,t})_{n,m} \;\equiv\; -\frac{\partial \log Q^d_{n,i,t}}{\partial p_{m,t}},$$

so diagonal elements are own-price elasticities and off-diagonal elements capture cross-asset substitution/complementarity. Under generalized isoelastic demand, investor i's log share demand is affine in the full log-price vector:

$$\log\left(\mathbf{Q}^d_{i,t}\right) = a_{i,t}\,\mathbf{1} + \mathbf{k}_{i,t} - \mathbf{E}_{i,t}\,\mathbf{p}_t, \tag{3}$$

where $\mathbf{Q}^d_{i,t} \in \mathbb{R}^N$ stacks $\{Q^d_{n,i,t}\}_{n=1}^N$, $\mathbf{p}_t \in \mathbb{R}^N$ stacks $\{p_{n,t}\}_{n=1}^N$, and $\mathbf{k}_{i,t} \in \mathbb{R}^N$ is a vector of non-price demand shifters.

We model these shifters as a function of observable asset characteristics,

$$\mathbf{k}_{i,t} = \mathbf{X}_t\,\boldsymbol{\beta}_{i,t} + \boldsymbol{\epsilon}_{i,t}, \tag{4}$$

where $\mathbf{X}_t \in \mathbb{R}^{N\times K}$ stacks asset characteristics (including a constant), $\boldsymbol{\beta}_{i,t} \in \mathbb{R}^K$ captures investor i's characteristic tilts, and $\boldsymbol{\epsilon}_{i,t} \in \mathbb{R}^N$ absorbs residual demand variation not explained by observables.

Market clearing Let $\mathbf{Q}^s_t \in \mathbb{R}^N$ denote exogenous shares outstanding. Aggregate demand is

$$\mathbf{Q}^d_t \equiv \sum_i \mathbf{Q}^d_{i,t},$$

and equilibrium prices satisfy market clearing,

$$\mathbf{Q}^d_t = \mathbf{Q}^s_t. \tag{5}$$

Equation (5) defines a high-dimensional fixed point in prices because demand depends on the full price vector through own- and cross-effects. Solving this fixed point exactly is generally infeasible at the stock-by-time frequency we need. We therefore work with a transparent first-order equilibrium approximation that preserves the two economic forces that matter for our counterfactuals: which investors are forced to trade (through their AUM shocks and ownership shares) and how easily the market absorbs these trades (through the aggregate demand elasticities).

Ownership shares, aggregate elasticities and capital reallocations. Let $Q^s_{n,t}$ denote shares outstanding. Investor i's ownership share in asset n is

$$s_{n,i,t} \equiv \frac{Q^d_{n,i,t}}{Q^s_{n,t}}, \qquad \mathbf{S}_{i,t} \equiv \mathrm{diag}\big(s_{1,i,t}, \ldots, s_{N,i,t}\big).$$

Ownership shares determine *who is marginal for each stock*: an investor with a larger $s_{n,i,t}$ has a larger mechanical impact on market clearing in asset n.

A first-order aggregation of market clearing implies that equilibrium price responses are governed by an *aggregate* elasticity matrix,

$$\mathbf{E}^{\mathrm{agg}}_t \equiv \sum_i \mathbf{S}_{i,t}\, \mathbf{E}_{i,t}, \qquad \mathbf{M}_t \equiv \big(\mathbf{E}^{\mathrm{agg}}_t\big)^{-1}, \tag{6}$$

where $\mathbf{E}_{i,t}$ is investor i's elasticity matrix from (3). We refer to $\mathbf{M}_t$ as the *multiplier matrix*. When cross-effects are small so that $\mathbf{E}^{\mathrm{agg}}_t$ is close to diagonal, $\mathbf{M}_t$ reduces to familiar stock-by-stock price multipliers. In general, off-diagonal elements of $\mathbf{M}_t$ capture equilibrium spillovers: forced trading in one set of stocks can propagate to others through substitution/complementarity.[1]

Our exercises reallocate capital across investor groups. Formally, the only primitive we shock is investor log AUM, $\{a_{i,t}\}$, while holding fixed the non-price demand shifters $\{\mathbf{k}_{i,t}\}$ and abstracting from endogenous AUM changes that would arise from the induced return changes. The key stock-level input is the *ownership-weighted reallocation exposure* vector,

$$\mathbf{f}_t \equiv \sum_i \mathbf{S}_{i,t}\, \Delta a_{i,t}\, \mathbf{1}, \tag{7}$$

which summarizes the mechanical buy/sell pressure created by the reallocation, asset by asset, given the pre-reallocation ownership shares.

[1]Appendix B.1 derives (6) from a first-order approximation to market clearing.

Proposition 1 (Equilibrium price impact) *Under generalized isoelastic demand and a first-order equilibrium approximation, a capital reallocation $\{\Delta a_{i,t}\}$ induces an equilibrium log-price change that can be written as*

$$\Delta \mathbf{p}_t = \underbrace{\mathbf{M}_t \, \mathbf{f}_t}_{\text{direct effect}} + \underbrace{\mathbf{M}_t \left(\sum_i \Delta \mathbf{S}_{i,t} \left(a_{i,t} \mathbf{1} + \mathbf{k}_{i,t} \right) \right)}_{\text{share re-composition}} + \underbrace{\Delta \mathbf{M}_t \left(\sum_i \mathbf{S}_{i,t} \left(a_{i,t} \mathbf{1} + \mathbf{k}_{i,t} \right) \right)}_{\text{multiplier change}}. \tag{8}$$

When the last two terms in (8) are quantitatively negligible, the price impact is

$$\Delta \mathbf{p}_t \approx \mathbf{M}_t \, \mathbf{f}_t. \tag{9}$$

The economic content of (9) is transparent. The *direct effect* depends only on (i) who is forced to trade, through the ownership-weighted reallocation exposure $\mathbf{f}_t$, and (ii) how easily the market absorbs this trading pressure, through the multiplier matrix $\mathbf{M}_t$. The two additional terms in (8) capture second-order adjustments: changes in ownership shares and changes in the aggregate multiplier. We validate in Section 3.2.3 that both are negligible, so focusing on (9) is both accurate and far more interpretable.

2.2 From Equilibrium Prices to Expected Returns: Pass-Through

A key point is that a price change generally does not translate one-for-one into expected returns. Even if a capital reallocation moves prices today, the adjustment need not be instantaneous: the price impact can fade only gradually as portfolios rebalance and capital reallocates over time. We discipline this mapping using the Campbell–Shiller present-value identity, following Davis et al. (2025).

Let $r_{n,t+1}$ denote the log return, $d_{n,t+1}$ the log dividend, and $p_{n,t}$ the log price. The Campbell–Shiller log-linearization implies

$$r_{n,t+1} = \kappa_n + (1 - \rho_n) \, d_{n,t+1} + \rho_n \, p_{n,t+1} - p_{n,t}, \tag{10}$$

where κ_n is a linearization constant and $\rho_n \in (0, 1)$ is the the price-to-(price+dividend) ratio.

We summarize price persistence by

$$\psi_n \;\equiv\; \frac{\partial\, \mathbb{E}_t[p_{n,t+1}]}{\partial p_{n,t}}, \tag{11}$$

and define $\mu_{n,t} \equiv \mathbb{E}_t[r_{n,t+1}]$.

Proposition 2 (Pass-through and expected-return rewriting) *Suppose an exogenous change in the current log price $p_{n,t}$ does not forecast next-period dividends,*

$$\frac{\partial\, \mathbb{E}_t[d_{n,t+1}]}{\partial p_{n,t}} = 0.$$

Then the expected-return response to an equilibrium log-price change satisfies

$$\Delta\mu_{n,t} \;\approx\; -\,\theta_n\, \Delta p_{n,t}, \qquad \theta_n \equiv 1 - \rho_n \psi_n, \tag{12}$$

where ψ_n is defined in (11). Stacking across assets and combining with Proposition 1 yields the mapping from capital reallocations to expected returns:

$$\Delta\boldsymbol{\mu}_t \;\approx\; -\,\mathbf{D}(\boldsymbol{\theta})\, \mathbf{M}_t\, \mathbf{f}_t, \tag{13}$$

where $\boldsymbol{\mu}_t \in \mathbb{R}^N$ stacks $\{\mu_{n,t}\}$, $\boldsymbol{\theta} \in \mathbb{R}^N$ stacks $\{\theta_n\}$, and $\mathbf{D}(\boldsymbol{\theta})$ is the diagonal matrix with θ_n on the diagonal.

A capital reallocation can push prices up or down today, but the expected-return effect depends on what the market expects prices to do next. If the price impact is mostly *transitory* (low ψ_n), a price drop today is expected to be followed by a rebound, so one-period expected returns rise nearly one-for-one with the initial price move. If the impact is mostly *persistent* (high ψ_n), there is little expected rebound, so the one-period expected-return response is muted. The pass-through coefficient θ_n summarizes this logic: pass-through is smaller when prices are more persistent (higher ψ_n) and when the asset is longer lived (higher ρ_n), so a given price change rewrites expected returns less.

Importantly, this step does not rely on any particular demand system. Unlike the equilibrium price-impact mapping, which depends on how investors substitute across assets, the price-to-expected-return mapping follows from the Campbell–Shiller present-value identity plus a mild identifying restriction: an exogenous change in price does not forecast dividends. Under this restriction, θ_n is pinned down by observable price dynamics, so the pass-through discipline is largely orthogonal to the specific demand structure used to generate the initial price impact.

2.3 From Stock-Level Expected Returns to Strategy Alpha and Capacity

We now aggregate the stock-level expected-return changes in (13) into portfolio outcomes. The previous section delivers how a capital reallocation rewrites *expected returns stock by stock*. This section answers the portfolio question: how does the same reallocation rewrite the *alpha* of a long–short strategy, and how much capital must be reallocated to eliminate that alpha?

A key discipline is that we treat reallocations as *within-market* shifts in delegated capital. Total assets under management are unchanged: dollar inflows to one investor group are funded by dollar outflows from another. We therefore parameterize each experiment as moving C dollars from a *donor* group D to a *recipient* group R, and we impose that each group scales its existing holdings proportionally (i.e., within-group portfolio weights are held fixed, as in Lou (2012)). Let $\alpha_{P,t}$ denote the alpha of strategy P at time t; the objects below map the reallocation size C into the implied change in $\alpha_{P,t}$ and into the corresponding strategy *capacity*.

Donor–recipient ownership overlap When C dollars move from a donor group D to a recipient group R, the same dollar reallocation does not create the same pressure in every stock. The reason is simple: donor and recipient groups can hold different portfolios. Stocks that are heavily owned by the donor group are forced to be sold; stocks that are heavily owned by the recipient group are forced to be bought. This *ownership overlap* is the key cross-sectional object that determines where the reallocation hits.

Appendix B.2 shows that when C dollars move from a donor group D to a recipient group R and inflows/outflows are distributed proportionally to investor size within each group, the ownership-weighted reallocation vector $\mathbf{f}_t$ is linear in C. It is convenient to separate the *cross-sectional trade list* from the *size* of the reallocation. We therefore write

$$\mathbf{f}_t \;\approx\; \frac{C}{G_t}\cdot \mathbf{g}_t, \qquad \mathbf{g}_t \equiv G_t\left(-\frac{1}{A_{D,t}}\sum_{i\in D}\mathbf{S}_{i,t}\mathbf{1} + \frac{1}{A_{R,t}}\sum_{i\in R}\mathbf{S}_{i,t}\mathbf{1}\right), \tag{14}$$

where

$$A_{D,t} \equiv \sum_{i\in D} A_{i,t}, \qquad A_{R,t} \equiv \sum_{i\in R} A_{i,t}$$

denote total AUM in the two groups. We refer to $\mathbf{g}_t$ as the *donor–recipient ownership overlap.*[2] The scalar G_t is a normalization that lets us express the reallocation size as a dimensionless intensity C/G_t while keeping $\mathbf{g}_t$ as the corresponding normalized trade list. Intuitively, $\mathbf{g}_t$ is the *direction* of the forced trades: it assigns positive weight to stocks that the recipient group overweights relative to the donor (net buys) and negative weight to stocks that the donor overweights (net sells), with magnitudes scaled by ownership. In our baseline implementation we set $G_t = A_{D,t}$, so C/G_t is simply the donor outflow fraction.

Alpha weights Let P denote a generic self-financing long–short strategy with long leg L and short leg S. Let $\mathbf{w}_{L,t}, \mathbf{w}_{S,t} \in \mathbb{R}^N$ and $\mathbf{w}_{S,t} \in \mathbb{R}^N$ denote the observed portfolio weights within each leg (summing to one within each leg), and let $\mathbf{w}_{M,t} \in \mathbb{R}^N$ denote market weights. Let β_S and β_L denote the strategy's market betas for each leg.

To translate stock-level expected-return changes into changes in $\alpha_{P,t}$, we collect the relevant weights in a single vector, which we refer to as the strategy's *alpha weights*:

$$\mathbf{h}_t \;\equiv\; \mathbf{w}_{L,t} - \mathbf{w}_{S,t} - (\beta_L - \beta_S)\,\mathbf{w}_{M,t}. \tag{15}$$

The interpretation is immediate. The first two terms capture the strategy's long-minus-short tilt across stocks. The last term strips out the part of the strategy that is simply market exposure. As a result, $\mathbf{h}_t$ isolates the cross-section that matters for *alpha rewriting*: a stock-

[2] Because $s_{n,i,t} = Q^d_{n,i,t}/Q^s_{n,t} = (w_{n,i,t}A_{i,t})/(P_{n,t}Q^s_{n,t}) = (w_{n,i,t}A_{i,t})/\mathrm{ME}_{n,t}$ with $\mathrm{ME}_{n,t} \equiv P_{n,t}Q^s_{n,t}$, we can write

$$g_{n,t} = \frac{G_t}{\mathrm{ME}_{n,t}}\left(\bar{w}_{n,R,t} - \bar{w}_{n,D,t}\right), \qquad \bar{w}_{n,R,t} \equiv \sum_{i\in R}\left(\frac{A_{i,t}}{A_{R,t}}\right)w_{n,i,t}, \qquad \bar{w}_{n,D,t} \equiv \sum_{i\in D}\left(\frac{A_{i,t}}{A_{D,t}}\right)w_{n,i,t}.$$

Thus $\mathbf{g}_t$ is the donor–recipient difference in group-level portfolio weights, scaled by inverse market equity and by the normalization G_t.

level expected-return change affects the strategy's alpha only to the extent that it loads on these alpha weights.

Proposition 3 (Alpha rewriting under a $D \to R$ capital reallocation) *Consider a reallocation that moves C dollars from a donor group D to a recipient group R, with investors in each group scaling their existing portfolios proportionally. Let $\mathbf{g}_t$ denote the donor–recipient ownership overlap in (14), and let $\mathbf{h}_t$ denote the strategy's alpha weights in (15). Under Propositions 1 and 2, the strategy's alpha change satisfies*

$$\Delta\alpha_{P,t} \;\approx\; -\frac{C}{G_t} \cdot \underbrace{\mathbf{h}_t' \, \mathbf{D}(\boldsymbol{\theta}) \, \mathbf{M}_t \, \mathbf{g}_t}_{\equiv \Lambda_{P,t}}, \tag{16}$$

where $\Lambda_{P,t}$ is the strategy's alpha sensitivity per unit of reallocation intensity *C/G_t.*

Equation (16) is an accounting of *who trades what* and *how prices and expected returns respond*. The donor–recipient overlap $\mathbf{g}_t$ is the normalized trade list induced by the reallocation. It is large when donor and recipient portfolios are segmented—their scaled weights differ materially—so the reallocation forces substantial net buying in some stocks and net selling in others, whether through a few concentrated trades or many dispersed ones. It is small when the two portfolios are similar and broadly diversified, so the induced trades are weak and diffuse. The multiplier matrix $\mathbf{M}_t$ converts this trade list into equilibrium price pressure, allowing for spillovers across assets through cross-substitution. The pass-through vector $\boldsymbol{\theta}$ converts equilibrium price pressure into changes in expected returns. Finally, the alpha weights $\mathbf{h}_t$ aggregate these stock-level expected-return changes into the strategy's alpha by focusing exactly on the long-minus-short component orthogonal to the market.

The single scalar $\Lambda_{P,t}$ therefore summarizes the entire channel: alpha erosion is large when the reallocation forces trading in the stocks that matter for alpha (alignment of $\mathbf{g}_t$ with $\mathbf{h}_t$), when demand is inelastic (large multipliers), and when price impacts translate strongly into expected returns (high pass-through).

Capacity We define strategy capacity $C^*_{P,t}$ as the dollar reallocation required to eliminate the observed (baseline) alpha $\bar{\alpha}_{P,t}$,

$$\bar{\alpha}_{P,t} + \Delta\alpha_{P,t} = 0.$$

Using (16), capacity is therefore

$$C^*_{P,t} = \left(\frac{G_t}{\Lambda_{P,t}}\right) \bar{\alpha}_{P,t}. \tag{17}$$

Capacity is the ratio of "how much alpha there is to erase" ($\bar{\alpha}_{P,t}$) to "how fast alpha erodes per dollar reallocated." Since $\Lambda_{P,t}$ is defined per unit of reallocation intensity C/G_t, the implied alpha-erosion rate per dollar is $\Lambda_{P,t}/G_t$. A high-capacity strategy either starts with a large baseline alpha or is hard to erase because reallocating capital has a small marginal effect on alpha (small $\Lambda_{P,t}$). A low-capacity strategy is the opposite: a modest reallocation produces a large alpha response, so $\Lambda_{P,t}$ is large and $C^*_{P,t}$ is small.

3 Estimating Components of Alpha Sensitivity $\Lambda_{P,t}$

3.1 Data Description

We combine CRSP, Compustat, and institutional holdings data to construct stock characteristics, investor classifications, and investor-by-stock ownership. Our stock sample includes all CRSP common stocks (share codes 10 and 11) on NYSE/AMEX/NASDAQ from January 1981 to December 2019, with accounting variables (e.g., book equity) from Compustat. For mutual funds, we start from the CRSP Survivor-Bias-Free Mutual Fund Database, consolidate share classes into fund entities using WRDS MFLINKS, and restrict to domestic equity funds that hold at least 80% in equities and exceed $15 million in average assets. We remove index funds using a name-based screen and then merge the remaining active funds to Morningstar Direct style categories, focusing on the nine standard size-by-style boxes (small/mid/large × value/blend/growth).

We classify all other equity holders into broad investor types (e.g., index funds, banks, insurance, advisors, pensions, endowments, households) following the Koijen–Yogo taxonomy. Holdings come from Thomson Reuters mutual fund holdings (s12) and institutional holdings (s34), with proper merging to avoid double counting mutual fund positions; household ownership is defined as residual shares outstanding net of all observed institutional holdings. Appendix A provides full construction details, screens, and classification steps.

3.2 Estimating Price Multipliers $\mathbf{M}_t$

3.2.1 Methodology

Our counterfactuals are governed by the multiplier matrix $\mathbf{M}_t = (\mathbf{E}_t^{\text{agg}})^{-1}$, where the aggregate elasticity matrix is the ownership-weighted average $\mathbf{E}_t^{\text{agg}} = \sum_i \mathbf{S}_{i,t}\mathbf{E}_{i,t}$. We therefore estimate investor-level elasticity matrices $\{\mathbf{E}_{i,t}\}$ and then construct $\mathbf{M}_t$ by aggregation and inversion. We estimate elasticities at the investor-group level defined in Section 3.1 and treat each group as a representative investor i.

A parsimonious cross-elasticity structure A fully unrestricted $N \times N$ matrix $\mathbf{E}_{i,t}$ is infeasible because it contains N^2 free parameters per investor and date, far exceeding the information available in a single cross-section of portfolio weights. We therefore impose a disciplined structure that still delivers a full cross-elasticity matrix: cross-price effects are stronger for stocks that are "close" substitutes, and they can differ within versus across industries as in Li et al. (2026). Concretely, let $g(n)$ denote asset n's industry and let $\mathbf{x}_{n,t} \in \mathbb{R}^K$ collect the stock characteristics that govern similarity. We measure how close two stocks n and m are by their Mahalanobis distance,

$$d_{n,m,t} \;\equiv\; \sqrt{(\mathbf{x}_{n,t} - \mathbf{x}_{m,t})'\,\boldsymbol{\Omega}_t^{-1}\,(\mathbf{x}_{n,t} - \mathbf{x}_{m,t})},$$

where $\boldsymbol{\Omega}_t$ is the cross-sectional covariance matrix of characteristics. This distance rescales each characteristic by its dispersion and accounts for correlations across characteristics, so "closeness" is measured in economically meaningful units rather than in raw levels.

We translate distance into substitution strength using a similarity kernel

$$K(d) \equiv \exp\big(-\tfrac{1}{2}d^2\big),$$

which satisfies $K(0) = 1$ and declines smoothly toward zero as d increases. Economically, this captures the idea that investors substitute most strongly across stocks that look similar on observables (e.g., similar size and B/M, profitability, investment), while cross-price effects between very dissimilar stocks are negligible.

We parameterize investor i's elasticity matrix as

$$(\mathbf{E}_{i,t})_{n,n} = \eta_{i,t}, \qquad (\mathbf{E}_{i,t})_{n,m} = \Big(\eta^{W}_{i,t}\, \mathbb{1}\{g(n) = g(m)\} + \eta^{A}_{i,t}\, \mathbb{1}\{g(n) \neq g(m)\}\Big)\, K(d_{n,m,t}), \;\; m \neq n, \tag{18}$$

so $\eta_{i,t}$ is the own-price elasticity and $(\eta^{W}_{i,t}, \eta^{A}_{i,t})$ govern within- and across-industry substitution among "similar" stocks. Economically, (18) makes the two central substitution margins explicit: investors can rotate *within* an industry (captured by $\eta^{W}_{i,t}$), and they can rotate *across* industries (captured by $\eta^{A}_{i,t}$) toward stocks with similar characteristics (captured by $K(d_{n,m,t})$). This parametrization ensures both margins are strongest among stocks that look most alike on observables.

From elasticities to an estimable regression We write our estimating equation directly in terms of portfolio weights. Under generalized isoelastic demand, log weights are affine in the full log-price vector:

$$\log w_{n,i,t} = k_{n,i,t} + p_{n,t} - (\mathbf{E}_{i,t}\mathbf{p}_t)_n.\text{[3]} \tag{19}$$

Given the elasticity structure in (18), define the within- and across-industry kernel-weighted price aggregates

$$Y^{W}_{n,t} \equiv \sum_{m \neq n:\, g(m)=g(n)} K(d_{n,m,t})\, p_{m,t}, \qquad Y^{A}_{n,t} \equiv \sum_{m \neq n:\, g(m)\neq g(n)} K(d_{n,m,t})\, p_{m,t}. \tag{20}$$

With these definitions,

$$(\mathbf{E}_{i,t}\mathbf{p}_t)_n = \eta_{i,t}\, p_{n,t} + \eta^{W}_{i,t}\, Y^{W}_{n,t} + \eta^{A}_{i,t}\, Y^{A}_{n,t},$$

so the demand system collapses to a single cross-sectional regression:

$$\log w_{n,i,t} = \underbrace{(1 - \eta_{i,t})\, p_{n,t}}_{\text{own-price}} \underbrace{-\, \eta^{W}_{i,t}\, Y^{W}_{n,t} \;-\; \eta^{A}_{i,t}\, Y^{A}_{n,t}}_{\text{cross-price}} + \underbrace{\boldsymbol{\gamma}'_{i,t}\mathbf{x}_{n,t} \;+\; u_{n,i,t}}_{\text{demand shifters}}, \tag{21}$$

[3] This follows from the generalized isoelastic log-share demand equation $\log Q^{d}_{n,i,t} = a_{i,t} + k_{n,i,t} - (\mathbf{E}_{i,t}\mathbf{p}_t)_n$ combined with the identity $Q^{d}_{n,i,t} = w_{n,i,t}A_{i,t}/P_{n,t}$, which implies $\log w_{n,i,t} = \log Q^{d}_{n,i,t} + p_{n,t} - a_{i,t}$.

estimated over inside assets with $w_{n,i,t} > 0$.[4] Equation (21) identifies $(\eta_{i,t}, \eta^W_{i,t}, \eta^A_{i,t})$. Together with (18), these estimates imply the full matrix $\widehat{\mathbf{E}}_{i,t}$ and therefore the multiplier matrix $\widehat{\mathbf{M}}_t = \left(\sum_i \mathbf{S}_{i,t}\widehat{\mathbf{E}}_{i,t}\right)^{-1}$.

Prices are endogenous because the latent demand component $u_{n,i,t}$ is generally correlated with equilibrium prices $p_{n,t}$, and therefore also with the constructed aggregates $(Y^W_{n,t}, Y^A_{n,t})$. To address this endogeneity, we instrument all price terms using a Koijen–Yogo style instrument that shifts demand through exogenous variation in investment universes. Let $p^{IV}_{n,t}$ denote the instrument for $p_{n,t}$. We then estimate (21) by 2SLS using instruments $(p^{IV}_{n,t}, Z^W_{n,t}, Z^A_{n,t})$, where $Z^W_{n,t}$ and $Z^A_{n,t}$ are the corresponding kernel-weighted instrument aggregates.[5]

3.2.2 Results

Panel A of Table 1 summarizes the estimated aggregate elasticity matrix. In each quarter, we compute cross-sectional percentiles, means, and medians across stocks for own- and cross-elasticity entries and then average these moments over time. Own-price elasticities are economically meaningful: across percentiles they range from 0.381 to 0.845, with a mean of 0.552. Cross-elasticities are orders of magnitude smaller, implying that substitution-driven spillovers across individual stocks are quantitatively limited in the aggregate.

Panel B reports the same statistics for the implied multiplier matrix, $\mathbf{M}_t = (\mathbf{E}^{\text{agg}}_t)^{-1}$. Own multipliers range from 1.663 to 2.968 across percentiles, with a mean of 2.339. The off-diagonal multipliers are again economically small, reinforcing the message from Panel A: the dominant channel is stock-level own-price pressure rather than propagation through cross-stock spillovers. The magnitude of the own multipliers is in line with the existing literature, which typically reports stock-level multipliers in the 2–4 range (e.g., Koijen and Yogo (2019); Gabaix and Koijen (2022)).

Substitution patterns across stocks strongly depend on the chosen cross-elasticity functional form. Haddad et al. (2025) show that with bilinear cross-elasticities, own and cross components are not separately identified without panel variation. Bilinear structure also

[4] We include a constant in $\mathbf{x}_{n,t}$, so we do not add a separate investor-specific intercept. We assign industries using the Fama–French 5-industry classification, following Li et al. (2026).

[5] Specifically,

$$Z^W_{n,t} \equiv \sum_{m\neq n:\, g(m)=g(n)} K(d_{n,m,t})\, p^{IV}_{m,t}, \qquad Z^A_{n,t} \equiv \sum_{m\neq n:\, g(m)\neq g(n)} K(d_{n,m,t})\, p^{IV}_{m,t}.$$

implies monotone substitutability in characteristics; for example, if a 5-year bond is a better substitute for another 5-year bond than a 1-year bond, monotonicity implies a 10-year bond must be an even better substitute, and a 20-year bond better still. By contrast, our cross-elasticity form is local, allowing assets with *closest* characteristics to be strongest substitutes rather than imposing global monotonicity. It also avoids the missing-intercept problem and can be identified in a cross section, without relying on the stronger exogeneity assumptions required for panel separation of own and cross effects.

He et al. (2025) emphasize that empirically measured demand elasticities in dynamic markets need not coincide with frictionless portfolio-choice benchmarks because prices move expectations and risk jointly. A central mechanism is volatility feedback: through the leverage effect (Black, 1976), price declines predict higher future volatility, which mechanically lowers desired holdings even for rational, unconstrained investors. The implication is that estimated elasticities are naturally *equilibrium* objects: they incorporate how conditional risk adjusts when prices move, not just a partial-equilibrium substitution slope holding risk fixed. This perspective fits our purpose. Our counterfactuals require the effective equilibrium multipliers that govern reallocation-driven repricing, so any volatility-feedback channel that dampens demand after price declines should be embedded in the $\mathbf{M}_t$ we estimate.

3.2.3 Validation

We validate the estimated multipliers by asking whether they match *real-world* repricing driven by demand shocks. Our counterfactuals map reallocation-induced trading pressure into equilibrium price changes. Fund flows provide a natural laboratory for this mapping: they generate large, observable shifts in investor demand that force trading through predetermined holdings. If the demand system captures the right equilibrium mechanics, the multiplier-implied repricing constructed from flows should align with reduced-form return effects.

The validation has two parts. First, in the full sample we construct flow-induced trading (FIT) from quarterly mutual-fund flows and lagged holdings, estimate the reduced-form return effect $\widehat{ret} \equiv \widehat{\gamma}\, FIT$ (with factor controls and quarter fixed effects), and compare it to the demand-system implied repricing $\widehat{\Delta p} \equiv \mathbf{M}FIT$. Second, we repeat the same comparison in an out-of-sample quasi-experiment: the June 2002 Morningstar rating methodology reform generates plausibly exogenous, rating-driven flows. Appendix Section D.1 reports the full exercise; we summarize the key evidence below.

Panel (a) of Figure 1 plots $\widehat{ret}$ against $\widehat{\Delta p}$ in the full sample (each point is a stock–quarter). The relation is tight and close to linear, with slope $\xi = 0.812$ and $R^2 = 0.95$, indicating that the estimated multipliers capture the cross-sectional magnitude of flow-driven repricing up to a modest scale factor. Panel (b) of Figure 1 repeats the same plot for the 2002Q3 Morningstar reform (each point is a stock). The fit is even tighter (slope $\xi = 0.806$, $R^2 = 0.99$), confirming that the mapping is not driven by typical flow episodes but also holds in quasi-experimental variation.[6]

To connect this stock-level validation to the portfolio-level objects used in our main analysis, we sort stocks in 2002Q3 into ten value-weighted book-to-market deciles and compute both the reduced-form rating-driven return effect and the demand-system implied repricing for each portfolio. Figure 2 shows two patterns: the two measures track each other closely across deciles, and repricing is economically intuitive. Value portfolios (high B/M) experience negative price pressure, while growth portfolios (low B/M) experience positive price pressure, reflecting reform-induced outflows from downgraded value funds and inflows into upgraded growth funds.[7]

[6]The scale of the demand shocks differs across panels. In the full sample, FIT reflects realized fund flows of all magnitudes, while in the Morningstar experiment the flow component is the predicted, rating-driven flow (e.g., the change associated with a one-star upgrade), which is mechanically smaller and therefore compresses the range of $\widehat{ret}$ and $\widehat{\Delta p}$. This difference is not important for our purpose because both axes scale linearly with the size of the demand shock; the validation is about proportionality (slope and fit), not about the shock magnitude.

[7]A useful way to read the flow-shock validation is as a calibration check for the magnitude of price impact implied by our estimated multipliers. In the validation regression, the slope ξ linking the reduced-form flow return effect to the demand-system implied repricing is slightly below one (about 0.8), which suggests that our baseline $\widehat{\Delta p}$ is modestly larger than the repricing implied by the flow laboratory. If one instead imposed the flow-based magnitude by rescaling multipliers so that $\widehat{\Delta p}$ matches the reduced-form effect (i.e., $\mathbf{M}_t^{\text{cal}} \approx \xi\,\mathbf{M}_t$), then all counterfactual price impacts and alpha responses would scale down proportionally. Since capacity is defined in Equation (17), and $\Lambda_{P,t}$ is linear in $\mathbf{M}_t$, capacity would mechanically scale up by $1/\xi$ (roughly 20–25%). Thus the flow-based calibration, if anything, pushes our capacity estimates upward and strengthens the conclusion that large reallocations are required to eliminate anomaly alphas.

3.3 Estimating Pass-Through $\boldsymbol{\theta}$

3.3.1 Methodology

We next estimate the pass-through coefficients $\boldsymbol{\theta}$ that map equilibrium log-price changes into changes in expected returns (Proposition 2). The object of interest is $\theta_n \equiv -\partial\mu_{n,t}/\partial p_{n,t}$: how much a one-unit increase in the current log price lowers next-period expected returns, holding cash-flow expectations fixed.

Let $P_{n,t}$ denote market equity and $BE_{n,t}$ book equity. Define $p_{n,t} \equiv \log P_{n,t}$, $b_{n,t} \equiv \log BE_{n,t}$, and $v_{n,t} \equiv p_{n,t} - b_{n,t}$ (log market-to-book). Let $\Delta p_{n,t} \equiv p_{n,t} - p_{n,t-1}$ be the log price change, and let $r_{n,t+1}$ denote the one-period-ahead log *total return* (including dividends). We work at an annual frequency, consistent with our annual estimation of anomaly α and β.

Proposition 2 identifies pass-through under the restriction that the price variation used for identification does not forecast future cash flows. We therefore isolate discount-rate-driven variation in prices using price wedges from an external discounted-cash-flow valuation in Van Binsbergen et al. (2023). Let $\hat{p}_{n,t}$ denote the corresponding log discounted-cash-flow price and define the wedge

$$PW_{n,t} \equiv p_{n,t} - \hat{p}_{n,t}, \qquad \Delta PW_{n,t} \equiv PW_{n,t} - PW_{n,t-1}.$$

Intuitively, $\hat{p}_{n,t}$ tracks the cash-flow component of valuation, so movements in $PW_{n,t}$ capture discount-rate and valuation forces rather than changes in expected cash flows.

We then estimate pass-through using a 2SLS design that instruments both the *price innovation* $\Delta p_{n,t-1}$ and the *valuation level* $v_{n,t-1}$. We include both because a price level shift affects expected returns through two distinct channels: the innovation Δp captures discount-rate *news* (how much valuation changes today), while the level v captures the *valuation state* (how persistent that valuation component is expected to be). Using both instruments lets us separately identify these "news" and "state" components of pass-through.

2SLS specification We estimate pass-through by instrumenting both the price innovation $\Delta p_{n,t-1}$ and the valuation level $v_{n,t-1}$ with two wedge-based instruments,

$$z^{\Delta}_{n,t} \equiv \Delta PW_{n,t} \quad \text{and} \quad z^{L}_{n,t} \equiv PW_{n,t},$$

which isolate discount-rate variation in price changes and in valuation levels, respectively. The first stages are

$$\Delta p_{n,t-1} = \pi_0 + \pi_1 z^{\Delta}_{n,t-1} + \pi_2 z^{L}_{n,t-1} + X'_{n,t-1}\pi_X + \xi^{\Delta}_{n,t-1}, \tag{22}$$

$$v_{n,t-1} = \kappa_0 + \kappa_1 z^{\Delta}_{n,t-1} + \kappa_2 z^{L}_{n,t-1} + X'_{n,t-1}\kappa_X + \xi^{v}_{n,t-1}, \tag{23}$$

where $X_{n,t-1}$ is a set of controls including investment-to-asset ratio, profitability, and market beta – consistent with the variables used in our demand-curve estimations. We use the same $X_{n,t-1}$ in the second stage,

$$r_{n,t} = a_0 + \beta_1\,\widehat{\Delta p}_{n,t-1} + \beta_2\,\widehat{v}_{n,t-1} + X'_{n,t-1}\gamma + \varepsilon_{n,t}. \tag{24}$$

The parameter we need is the response of expected returns to the current log price, $\theta_n \equiv -\partial\mu_{n,t-1}/\partial p_{n,t-1}$. In (24), a one-unit increase in $p_{n,t-1}$ shifts both regressors, $\Delta p_{n,t-1}$ and $v_{n,t-1}$, by one (holding $p_{n,t-2}$ and $b_{n,t-1}$ fixed). The combined effect on expected returns is therefore $\beta_1 + \beta_2$, so we estimate pass-through as

$$\widehat{\theta} \;\equiv\; -\big(\widehat{\beta}_1 + \widehat{\beta}_2\big). \tag{25}$$

3.3.2 Results

Table 2 reports 2SLS estimates of the pass-through parameter θ using price-wedge instruments, both in the pooled stock-level panel and for portfolios sorted by book-to-market and prior (2–12 month) return.

Panel A shows that the instruments are strong. They are highly relevant for both endogenous regressors, with most first-stage t-statistics in double digits. The Kleibergen–Paap rk Wald F statistics are large and comfortably above standard weak-instrument thresholds in all cases, with the growth portfolio the lone borderline case. Overall, the first stage supports a well-identified design.

Panel B reports the second-stage coefficients and the implied pass-through, $\widehat{\theta} = -\big(\widehat{\beta}_{\Delta p} + \widehat{\beta}_v\big)$. In the pooled panel, $\widehat{\theta} = 0.042$ and is statistically significant. For book-to-market portfolios, estimates are similar, ranging from 0.052 to 0.064, albeit with less precision due to aggregation. In contrast, pass-through is substantially larger for momentum portfolios: $\widehat{\theta} =$

0.100 for winners and $\widehat{\theta} = 0.458$ for losers (the latter estimated less precisely). Overall, pass-through for individual stocks and value/growth portfolios lies in the 4–6% range, consistent with highly persistent valuation components. Momentum portfolios exhibit much larger pass-through, consistent with the view that momentum is driven by more transitory expected-return variation and therefore translates price pressure into expected returns more strongly.

We use the pooled stock-level estimate as our baseline θ and apply it uniformly across all reallocations and portfolios. We prefer it because it is the most precisely estimated object and the portfolio-specific estimates, while broadly similar in magnitude, are noisier and would add sampling variation without changing the economic conclusions.

Van Binsbergen et al. (2025) emphasize a basic identification problem in forward-looking markets: the quantity response to a price change depends on what the underlying shock does to expectations. An instrument that moves prices today can simultaneously move expected returns and expected cash flows, so the observed portfolio adjustment is not a shock-invariant "demand slope"; different shocks trace out different elasticity objects. This critique is sharpest for elasticity concepts tied to *highly transitory* price movements, where a price dislocation is assumed to revert quickly and expected returns move almost one-for-one with the initial price change. Our framework is aligned with this perspective because we do not rely on a one-for-one pass-through elasticity. Instead, we separate the mapping from reallocations to equilibrium *prices* (via $\mathbf{M}_t$) from the mapping from prices to expected returns (via $\boldsymbol{\theta}$). With $\widehat{\theta}$ in the single-digit % range across the specifications, expected returns respond only partially to price pressure, so our counterfactuals naturally target the effects of more *persistent* reallocations across investor groups rather than temporary price dislocations.

3.4 Estimating Anomaly α and β

We estimate alphas and market betas for value and momentum using standard decile portfolios. For value, we form book-to-market (B/M) deciles each June following Fama and French (1992). Low B/M firms are "growth" stocks, while high B/M firms are "value" stocks. The value strategy goes long the highest-B/M decile (the *value* leg) and short the lowest-B/M decile (the *growth* leg). For momentum, we form deciles based on past returns following Jegadeesh and Titman (1993); the strategy goes long the top decile of past returns (the *winner* leg) and short the bottom decile (the *loser* leg). We compute annual (one-year) returns for each decile over the subsequent year.

To estimate market exposure and alpha, we run the CAPM time-series regression at an annual frequency for each decile portfolio i:

$$R_{i,t} - R_{TB,t} = \alpha_i + \beta_i\big(R_{m,t} - R_{TB,t}\big) + \varepsilon_{i,t}, \tag{26}$$

where $R_{i,t}$ is the decile portfolio's return, $R_{m,t}$ is the market return, and $R_{TB,t}$ is the Treasury bill rate. The intercept α_i is the portfolio alpha and β_i is the market beta. In the main analysis, we use the estimated α and β for the long leg, the short leg, and the corresponding long–short strategy.

Panel A of Table 3 reports average annual returns, CAPM alphas, and market betas for book-to-market (B/M) deciles. The return spread between the highest-B/M (value) and lowest-B/M (growth) portfolios is about 4.8% per year. The corresponding H–L alpha is about 5.6%, reflecting that growth stocks have a slightly higher market beta than value stocks. In other words, the value premium remains strong even after adjusting for market risk, consistent with the existing literature.

Panel B reports a momentum premium of about 13.6% per year. Betas differ sharply across legs: losers have a beta of 1.63 versus 1.21 for winners, a 0.42 gap. This beta differential mechanically boosts the long–short CAPM alpha, because the short leg loads more on the market and therefore gives up more expected return when we net out market risk.

4 Effects of Reallocations on Alpha

4.1 Reallocation scenarios

We study two reallocations that are both economically realistic and tightly linked to how investors gain exposure to the value anomaly. The first is *style rotation* within delegated management: investors move capital from growth funds to value funds. The second is a *delegation shift*: households move capital from direct stockholding into value funds. These experiments are useful for two reasons. First, they correspond to common investment decisions that market participants can observe and implement (e.g., reallocating across Morningstar style funds). Second, they isolate the key economic margin in our framework: *the source of capital*. Even when the recipient is the same (value funds), selling pressure looks very different when dollars come from growth funds versus from households.

Style rotation: Growth-to-value reallocation A growth-to-value reallocation models a standard style-rotation decision: investors reduce exposure to growth-oriented delegated portfolios and increase exposure to value-oriented delegated portfolios. This is a natural setting because Morningstar style classifications are widely used in practice and strongly align with the conventional value definition based on book-to-market. Panel (a) of Figure 3 confirms this alignment: across small, mid, and large size buckets, Morningstar value funds hold portfolios with substantially higher holdings-weighted B/M ratios than Morningstar growth funds. This makes growth-to-value flows a realistic and transparent way to inject capital into the value trade.

Importantly, style categories are not mechanically segmented in the data. Growth and value funds overlap in their holdings, so a growth-to-value rotation generates crisscross trades at the stock level. For example, value funds hold 4.43% of top-decile (value) stocks and 1.03% of bottom-decile (growth) stocks, while growth funds hold 5.54% of bottom-decile stocks and 1.48% of top-decile stocks. Our framework does not impose that growth funds only sell growth stocks and value funds only buy value stocks. Instead, the net reallocation pressure is determined by observed ownership at the stock level, so overlap is handled mechanically rather than by assumption.

Finally, growth-to-value rotation naturally generates *cross-anomaly* effects. Momentum is implemented by going long recent winners and short recent losers. Panel (b) of Figure 3 shows that growth funds tilt toward winners, while value funds tilt toward losers. As a result, reallocating capital from growth to value can move not only the value premium but also the momentum premium through overlapping holdings.[8] This is precisely the type of spillover our stock-level mapping is designed to capture.

Delegation shift: Households-to-value reallocation A households-to-value reallocation models a different but equally realistic channel for investing in value: households reduce direct stockholding and allocate more capital to delegated value funds. The key distinction is the *source* of dollars. Household portfolios span both growth and value stocks, so the

[8] In our sample, the correlation between the top-minus-bottom decile value and momentum portfolios is -0.203.

implied selling pressure is broad-based rather than concentrated in growth-heavy holdings. This makes the equilibrium repricing meaningfully different from a pure style rotation, even though the recipient is the same.

Studying households-to-value reallocations therefore isolates a core prediction of the framework: identical inflows into value funds can have different effects on prices and alphas depending on where the dollars come from. In the empirical results, we use this comparison to quantify how source-of-capital shape both the targeted impact on the value strategy and the spillover impact on momentum.

4.2 Results: reallocations and alpha rewriting

Table 4 reports how a large reallocation into value funds rewrites the value and momentum long–short strategies. In both experiments, the inflow into value funds is fixed at the same dollar amount—equal to 100% of aggregate growth-fund AUM.

The first row collects the baseline unconditional returns and CAPM alphas from Table 3. The next rows report the reallocation effects. When growth funds are the donor, the value strategy falls by about 0.74 percentage points in both expected return and CAPM alpha. The near one-for-one movement of $\mathbb{E}[r^{H-L}]$ and α^{CAPM} indicates that the reallocation primarily compresses the value spread through cross-sectional repricing rather than through changes in aggregate market pricing.[9] When households are the donor, the erosion is materially smaller: value expected returns fall by 0.41 percentage points and value alpha falls by 0.42. The same dollars flow into value funds, but the financing comes from a broad household portfolio, so selling pressure is diluted across the cross-section instead of being concentrated in the stocks that define the growth-versus-value spread.

Even at this extreme scale, value remains hard to erase. After reallocating an amount equal to 100% of growth-fund AUM, value alpha is still 4.86% when growth funds are the donor and 5.17% when households are the donor. Put differently, reallocations into value funds erode the value premium only gradually, and the origin of the dollars matters as much as the destination.

The same reallocation also generates a clear spillover to momentum. When growth funds are the donor, momentum expected returns and CAPM alpha both increase by about 0.41

[9]We compute alpha changes holding market betas fixed at their pre-reallocation estimates.

percentage points. This sign is consistent with overlap in holdings: the outflow is tilted toward stocks that load on the momentum winner leg, so the induced trade list steepens the winner-minus-loser spread going forward. When households are the donor, the spillover is much weaker (0.09 percentage points in expected returns and 0.06 in alpha), reflecting that household outflows are not particularly "winner-heavy" and therefore load less on momentum's long and short legs.

Finally, the capacity calculations summarize the same message in dollars. Driving the value CAPM alpha to zero would require about \$3.1 trillion when growth funds finance the reallocation (896% of growth-fund AUM) and about \$5.0 trillion when households finance it (94% of household equity AUM). Even very large reallocations into value funds therefore leave substantial value alpha intact, and the effectiveness depends sharply on where the dollars come from.

A final caveat is that our capacity calculations use a first-order mapping in which price impact is approximately linear in the reallocation size C. If instead price impact is *concave* in C—so that marginal price impact declines as shocks get larger—then our linear modelling will overstate alpha erosion for large reallocations and therefore understate capacity. Intuitively, concavity means that large reallocations are absorbed more easily than implied by a constant multiplier, so the same additional dollar moved produces a smaller incremental price change and a smaller incremental change in expected returns. As a result, the true dollar reallocation required to drive α to zero would be larger than our linear C^*, particularly in high-capacity scenarios. This logic is consistent with evidence in Chaudhry and Li (2025), who find that price impact is smaller for larger shocks; if anything, such concavity strengthens our conclusion that alphas are hard to eliminate via large-scale capital reallocation.

4.3 Capacity is state dependent

Figure 4 shows that value capacity is highly time-varying. Panel (a) plots the dollar reallocation required to eliminate value alpha under both reallocation scenarios. Capacity is small in the early 1980s, rises sharply from the mid-1990s onward, and exhibits large swings around major market episodes. Across the sample, the households-to-value capacity consistently exceeds growth-to-value capacity in dollar terms. This difference reflects that an identical inflow into value funds can trigger very different market-wide repricing depending on the initial portfolio of the funding source that is liquidated.

Panel (b) rescales the same object by donor AUM, $C^{*}_{P,t}/A_{D,t}$, and makes the source-of-capital margin even starker. Because growth-fund AUM is small early in the sample, the growth-to-value capacity ratio is extremely large in the 1980s and remains well above one throughout, implying that eliminating value alpha would require moving multiples of the donor base even in later decades. In contrast, the household-to-value capacity ratio is much smaller because households are a large donor base: even trillion-dollar capacities translate into modest fractions of household equity AUM.

Figure 5 explains *why* capacity moves over time by unpacking the alpha-sensitivity term $\Lambda_{P,t}$ from Proposition 3. Panel (a) shows that $\Lambda_{P,t}$ is highly time-varying in both reallocation scenarios.[10] The growth→value series exhibits larger fluctuations than the households→value series, consistent with style rotation at times becoming unusually well—or poorly—aligned with the value spread. The growth→value $\Lambda_{P,t}$ rises sharply in the late 1990s and early 2000s, spikes again around the global financial crisis, and declines after 2010. As the shares held by growth funds, value funds, and households evolve, the same reallocation intensity maps into a different ownership-weighted trade list (examined next) and therefore into a different alpha sensitivity.

Panel (b) isolates the first-order "who-trades-what" channel by plotting the raw alignment $\mathbf{h}_t'\mathbf{g}_t$. This object is large when the reallocation is both *large in effective terms* and *well targeted*: the donor base is sufficiently important in the ownership of the relevant stocks, and the induced trade list loads on the stocks that matter for value alpha. Two patterns stand out. First, the growth→value alignment rises over time and spikes around the late-1990s/early-2000s episode, consistent with periods when style portfolios were unusually polarized and style rotation forced trades directly into the value long leg and out of the growth short leg. Second, the households→value alignment follows a different path because households finance the inflow by selling a broad slice of the market; even when the dollar amount is large, the induced trade list is less concentrated in the stocks that define the value spread, so alignment is weaker.

Table 5 clarifies the ownership forces behind these movements by showing how growth

[10] The levels differ because $\Lambda_{P,t}$ is defined per unit of donor outflow fraction when we set $G_t = A_{D,t}$, so a unit reallocation intensity corresponds to different dollar magnitudes across donors.

funds, value funds, and households allocate across the B/M cross-section. Growth and value funds are tilted in opposite directions, but both hold a wide swath of stocks, and this breadth expands sharply over time (the "percentage of stocks" held rises across deciles in later decades). As delegated portfolios become broader and more overlapping, a style rotation becomes less of a clean "sell growth, buy value" trade at the stock level. Mechanically, this reduces the donor–recipient contrast embedded in $\mathbf{g}_t$ and weakens $\mathbf{h}_t'\mathbf{g}_t$, pushing capacity up all else equal.

Panel C adds the key benchmark for the households-to-value scenario: households are close to the market portfolio in B/M space. They hold essentially the entire cross-section (near-universal "percentage of stocks" across deciles), and their dollar-weight shares are far less tilted than delegated style funds. This is exactly why households-to-value reallocations generate weaker alignment: funding the same inflow into value funds requires selling a broad slice of the market rather than disproportionately selling the growth leg. In short, time variation in capacity reflects both the evolving segmentation of delegated style investing and the fact that households remain a broad, market-like donor base.

Panel (c) of Figure 5 isolates the equilibrium amplification channel by plotting $\left(\mathbf{h}_t'\mathbf{M}_t\mathbf{g}_t\right)/\left(\mathbf{h}_t'\mathbf{g}_t\right)$. This ratio answers a clean question: holding the trade list fixed, how much does the demand system amplify its effect in the direction relevant for value. The amplification factor is strongly time-varying but moves similarly across the two donor scenarios, as expected if it is mainly governed by market-wide demand conditions (through $\mathbf{M}_t$) rather than by the identity of the marginal seller. Importantly, amplification declines markedly over the sample: for the growth→value scenario it falls from roughly 2.8–2.9 in the early 1980s to about 1.3–1.4 by the mid-2010s (with a modest rebound afterward). Economically, this means that a given "raw" alignment $\mathbf{h}_t'\mathbf{g}_t$ translated into roughly twice as much equilibrium repricing early in the sample as it does today, which mechanically pushes capacity up over time.

Putting the results together clarifies what moves capacity. Differences *across* donor scenarios are primarily about the trade list and the size of the donor base. The donor-recipient pair determines whether the reallocation hits the value spread directly or gets diversified away, while the size of the donor group governs how much effective trading pressure a given reallocation intensity generates. Time variation *within* each scenario reflects two forces: slow-moving shifts in ownership and investor composition that change $\mathbf{g}_t$ over time, and changes in equilibrium amplification that determine how strongly a given trade list translates into prices.

5 Conclusion

This paper studies a simple but under-quantified equilibrium question: how much capital must reallocate to eliminate the alpha of an investment strategy, and what other alphas move along the way? Using a demand-based equilibrium mapping, we translate source–destination reallocations into stock-level price pressure, present-value pass-through into expected returns and strategy-level alpha erosion. The main empirical lesson is stark: reallocations can generate meaningful price impact yet translate into only modest alpha compression, and—because strategies overlap—reallocation can mechanically shift alpha across trades rather than eliminate it.

The first implication concerns how we interpret the persistence of prominent strategies. In an inelastic market, price pressure is not the bottleneck; *targeted* ownership shifts and pass-through are. This suggests a new way to read "failed arbitrage" narratives: alphas can survive even in the presence of large, sophisticated capital if the capital does not originate from the right investors, does not hit the right stocks, or produces repricing with limited expected-return content. In that sense, persistence is less a paradox and more an equilibrium accounting identity once one takes investor heterogeneity and ownership seriously.

The second implication is for policy and product design in an era of systematic investing. Strategy crowding is inherently a *system* problem: reallocations intended to compress one spread can mechanically inflate another through shared holdings, creating whack-a-mole dynamics and unintended spillovers. Stress-testing new products or allocation rules therefore requires an equilibrium lens that tracks source-of-capital constraints and overlap, not just stand-alone backtests. More broadly, the framework provides a practical tool for assessing when the next dollar of reallocation is likely to rewrite alpha —and where in the cross-section that rewriting will show up.

References

Asness, Clifford S., 1994, *Variables That Explain Stock Returns*, Ph.D. thesis, University of Chicago, Ph.D. Dissertation.

Barberis, Nicholas, and Andrei Shleifer, 2003, Style investing, *Journal of financial Economics* 68, 161–199.

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

Berk, Jonathan, and Richard Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Black, Fischer, 1976, Studies of stock price volatility changes, in *Proceedings from the American statistical association, business and economic statistics section*, 177.

Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *The Review of Financial Studies* 1, 195–228.

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, 2015, Regression discontinuity and the price effects of stock market indexing, *The Review of Financial Studies* 28, 212–246.

Chaudhary, Manav, Zhiyu Fu, and Jian Li, 2025, Corporate bond multipliers: Substitutes matter, *Working Paper* .

Chaudhry, Aditya, and Jiacui Li, 2025, Endogenous elasticities: Price multipliers are smaller for larger demand shocks, *Working Paper* .

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

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

Dannhauser, Caitlin D, and Jeffrey Pontiff, 2024, Flow, *Available at SSRN 3428702* .

Davis, Carter, 2024, Elasticity of quantitative investment, *Review of Financial Studies, Forthcoming* .

Davis, Carter, 2025, Classical versus demand system asset pricing, *Available at SSRN 5232345* .

Davis, Carter, Mahyar Kargar, and Jiacui Li, 2025, Why do portfolio choice models predict inelastic demand?, *Journal of Financial Economics* 172, 104096.

Evans, Richard B, 2010, Mutual fund incubation, *Journal of Finance* 65, 1581–1611.

Fama, Eugene F, and Kenneth R French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance* 47, 427–465.

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

Fuchs, William, Satoshi Fukuda, and Daniel Neuhann, 2023, Demand-system asset pricing: Theoretical foundations, *Available at SSRN 4672473* .

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

Haddad, Valentin, Zhiguo He, Paul Huebner, Péter Kondor, and Erik Loualiche, 2025, Causal inference for asset pricing, *Available at SSRN 5187305* .

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

He, Zhiguo, Peter Kondor, and Jessica S Li, 2025, Demand elasticity in dynamic asset pricing, *Available at SSRN 5297190* .

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance* 48, 65–91.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2008, Unobserved actions of mutual funds, *The Review of Financial Studies* 21, 2379–2416.

Koijen, Ralph S. J., and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Koijen, Ralph SJ, Robert J Richmond, and Motohiro Yogo, 2023, Which investors matter for equity valuations and expected returns?, *Review of Economic Studies* 1–38.

Korajczyk, Robert A, and Ronnie Sadka, 2004, Are momentum profits robust to trading costs?, *The Journal of Finance* 59, 1039–1082.

Li, Jiacui, and Zihan Lin, 2025, Price multipliers are larger at more aggregate levels, *Working Paper* .

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

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

Pástor, Luboš, and Robert F Stambaugh, 2012, On the size of the active management industry, *Journal of Political Economy* 120, 740–781.

Stambaugh, Robert F, 2014, Presidential address: Investment noise and trends, *The Journal of Finance* 69, 1415–1453.

Van Binsbergen, Jules H, Martijn Boons, Christian C Opp, and Andrea Tamoni, 2023, Dynamic asset (mis) pricing: Build-up versus resolution anomalies, *Journal of Financial Economics* 147, 406–431.

Van Binsbergen, Jules H, Benjamin David, and Christian C Opp, 2025, How (not) to identify demand elasticities in dynamic asset markets, Technical report, National Bureau of Economic Research.

Van der Beck, Philippe, 2025, Short- versus long-run demand elasticities in asset pricing, *Working paper* .

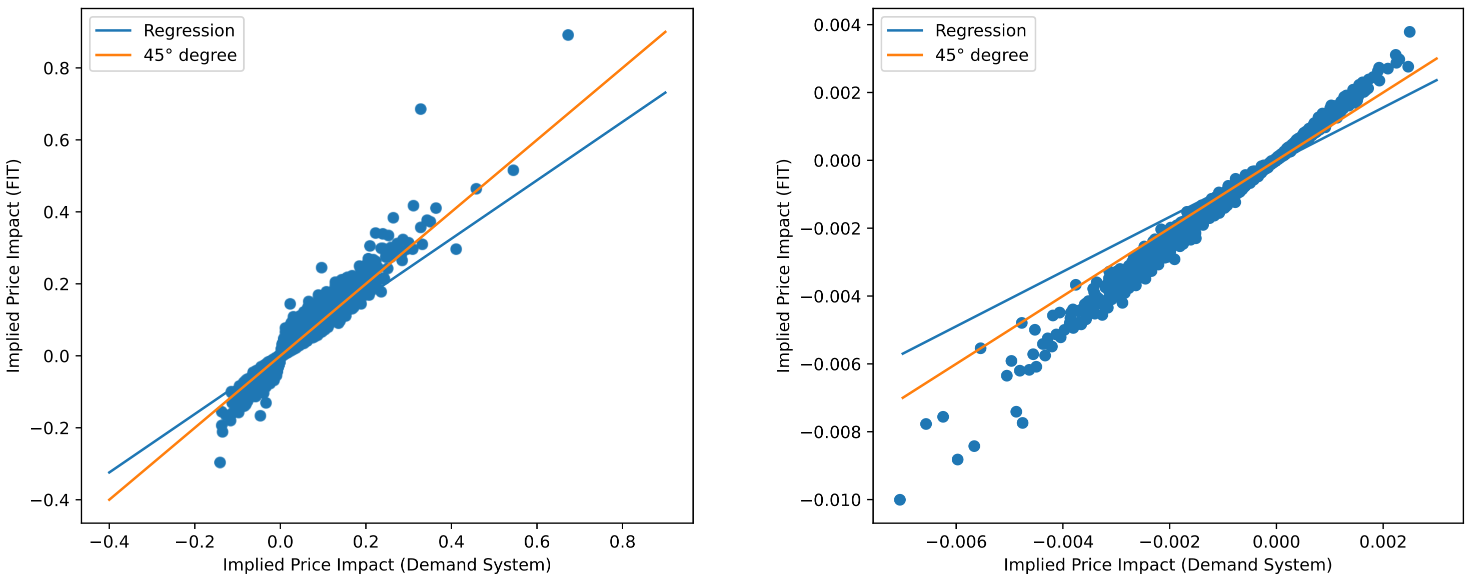
Van der Beck, Philippe, Jean-Philippe Bouchaud, and Dario Villamaina, 2025, Self-inflated funds, *Working paper* .

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

Watson, Towers, 2015, The world's 300 largest pension funds: Year end 2014.

Tables and Figures

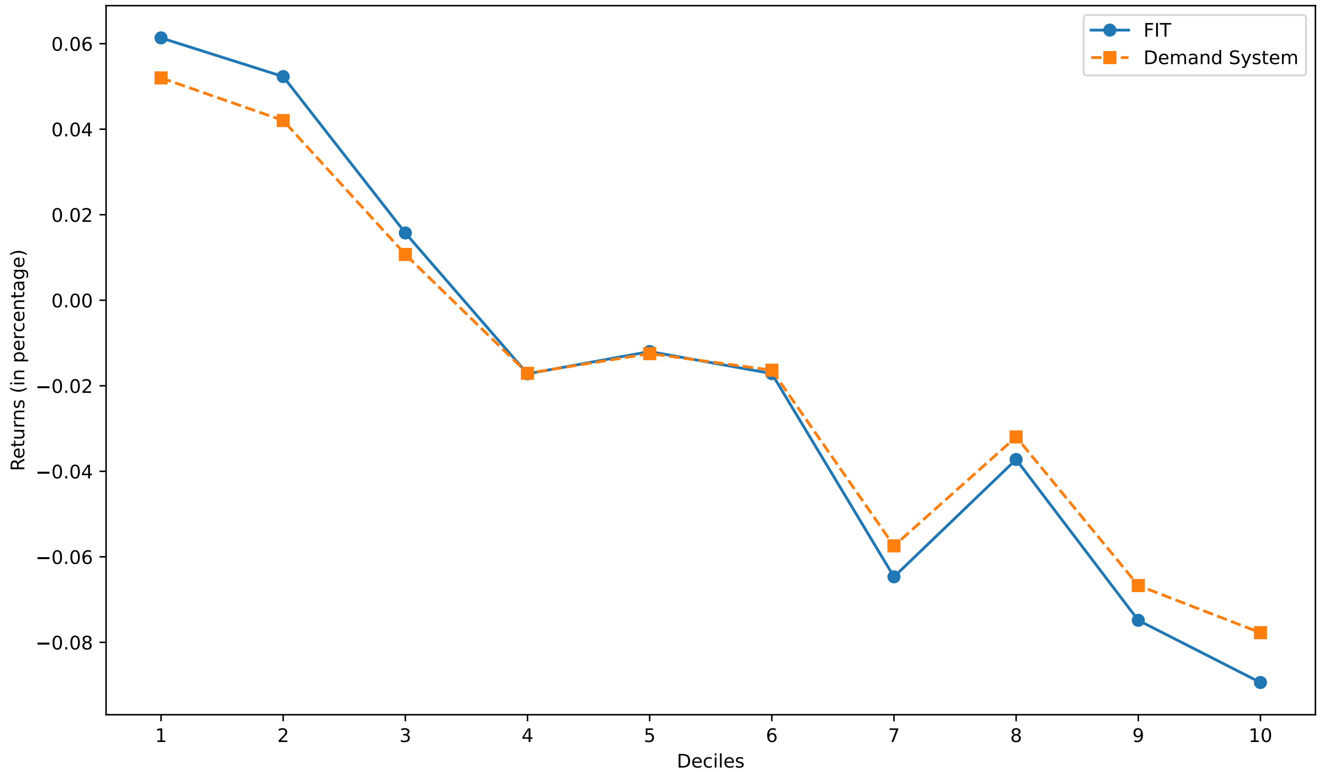
Figure 1: Validation of estimated price multipliers against reduced-form flow effects



(a) Full sample (1981–2019) **(b) Morningstar rating reform (2002Q3)**

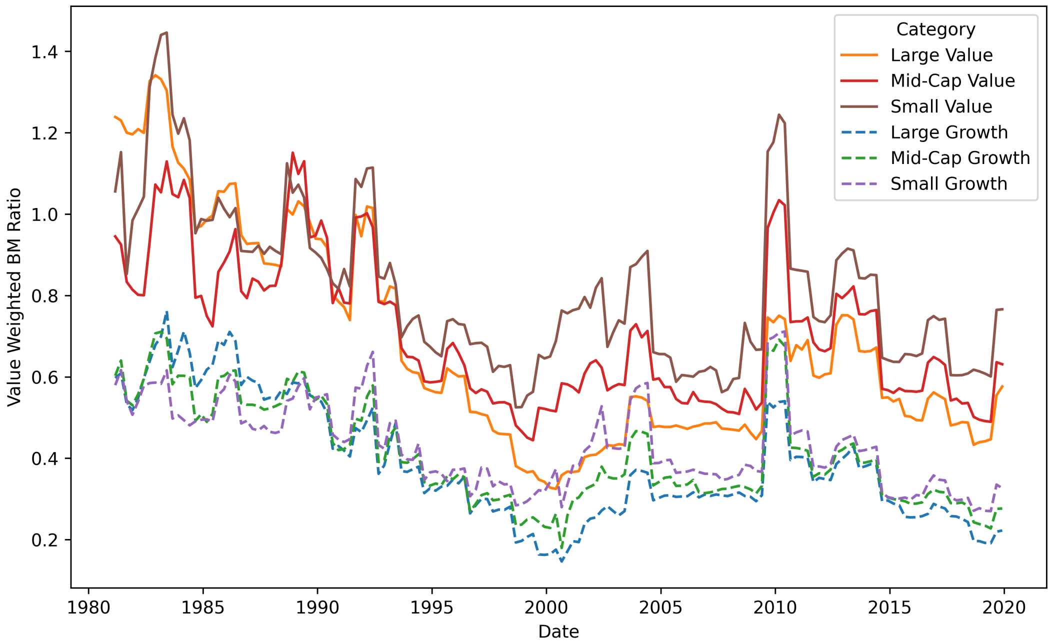
Both panels compare reduced-form return effects of flow-induced trading with demand-system implied log-price changes constructed from the estimated multipliers. In each panel, the horizontal axis is $\widehat{\Delta p}$, the model-implied log-price change from the demand system, and the vertical axis is $\widehat{ret}$, the reduced-form return effect implied by flow exposure. Specifically, $FIT = \sum_j s_{i,j}\, Flow_j$ aggregates fund-level flow shocks into stock-level exposure using lagged holdings, and $\widehat{ret} \equiv \widehat{\gamma}\, FIT$ uses the estimated return-on-FIT slope $\widehat{\gamma}$ (from the full sample with factor controls and time fixed effects). The model-implied repricing is $\widehat{\Delta p} \equiv \widehat{m}\, FIT$, where $\widehat{m}$ is the estimated stock-level multiplier. Panel (a) uses realized quarterly mutual-fund flows in the full sample; each point is a stock–quarter observation. Panel (b) uses quasi-experimental rating-driven flows from the June 2002 Morningstar methodology reform; each point is a stock in 2002Q3. A tight relation with slope close to one in both panels indicates that the estimated multipliers capture the magnitude of flow-driven repricing in the data.

Figure 2: Morningstar rating reform: portfolio-level repricing by book-to-market decile (2002Q3)

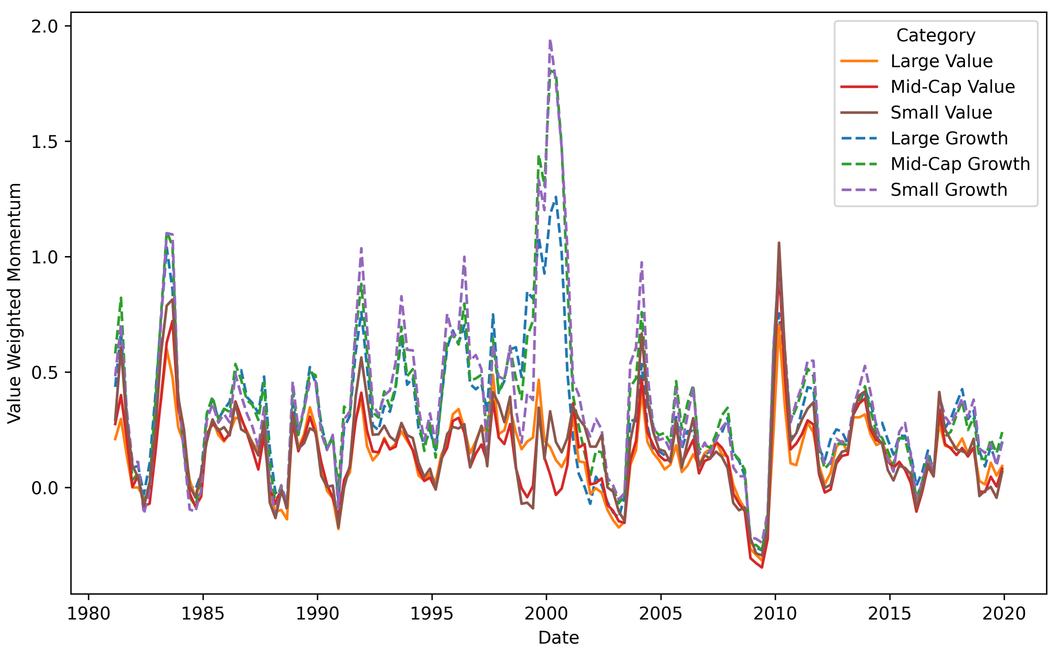


This figure aggregates the Morningstar 2002 rating-reform shock to the portfolio level. Stocks in 2002Q3 are sorted into ten value-weighted book-to-market (B/M) deciles using end-of-Q2 characteristics. For each decile, the figure plots (i) the reduced-form return effect implied by rating-driven flows, $\widehat{ret} \equiv \hat{\gamma}\, FIT$, where $FIT = \sum_j s_{i,j} \widehat{Flow}_j$, and (ii) the demand-system implied log-price change, $\widehat{\Delta p} \equiv \hat{m}\, FIT$, constructed from the estimated multipliers. Close alignment across deciles indicates that the estimated multipliers capture the cross-sectional pattern of flow-driven repricing induced by the quasi-experimental rating shock; in particular, the reform reallocates capital away from value funds and toward growth funds, generating negative repricing for high B/M stocks and positive repricing for low B/M stocks.

Figure 3: Style portfolios: holdings-weighted book-to-market and momentum profiles over time



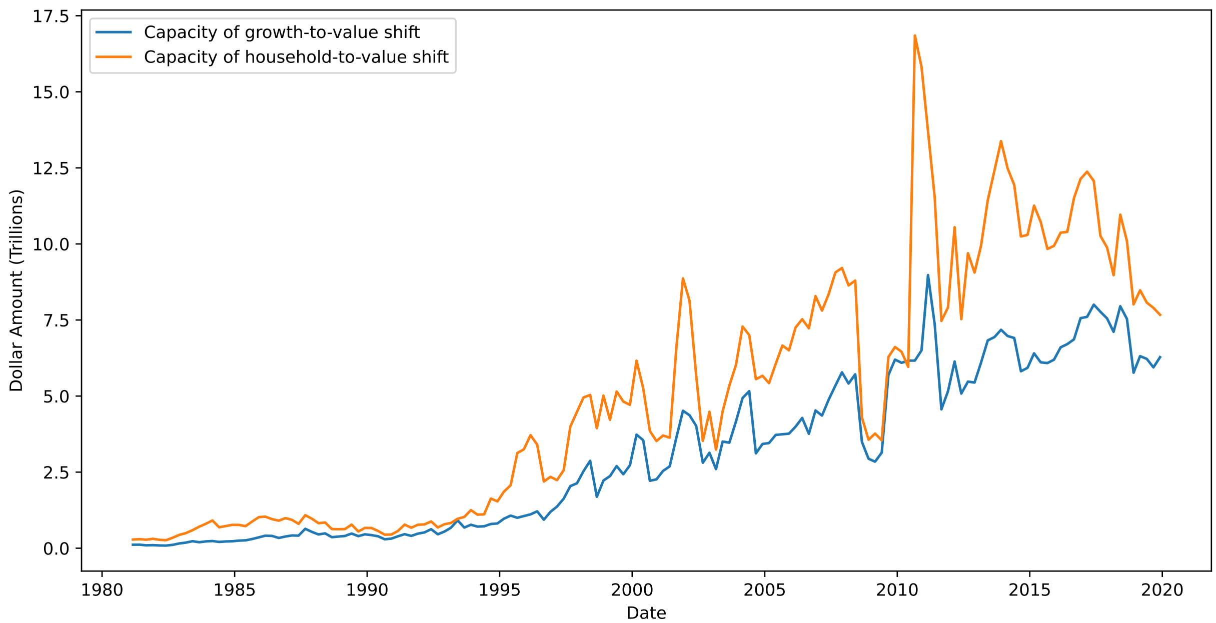
Panel (a): Holdings-weighted book-to-market by style and size



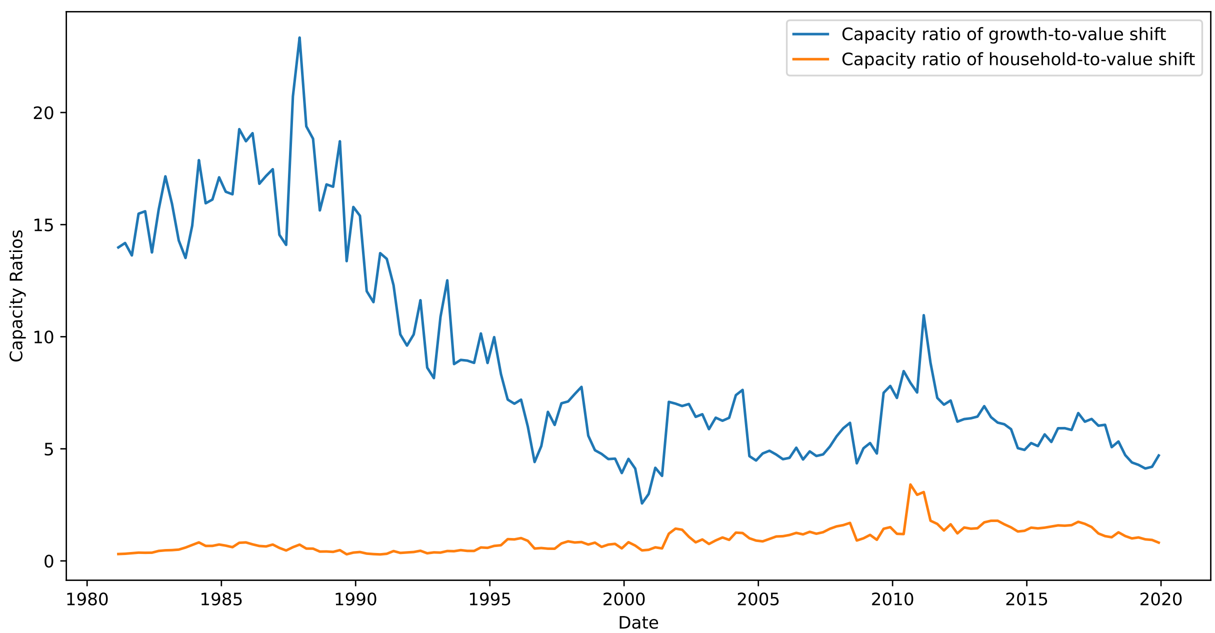
Panel (b): Holdings-weighted momentum by style and size

This figure summarizes how the six size-by-style active-fund categories tilt toward value and momentum over time. Panel A plots the holdings-weighted average book-to-market (B/M) ratio for Small-Value, Mid-Value, Large-Value, Small-Growth, Mid-Growth, and Large-Growth funds. Panel B plots the holdings-weighted average momentum score, measured as cumulative returns from month $t-12$ to $t-2$, for the same categories. In both panels, weights are based on funds holdings, so each series reflects the average characteristics of the stocks held by that fund category.

Figure 4: Time variation in value capacity



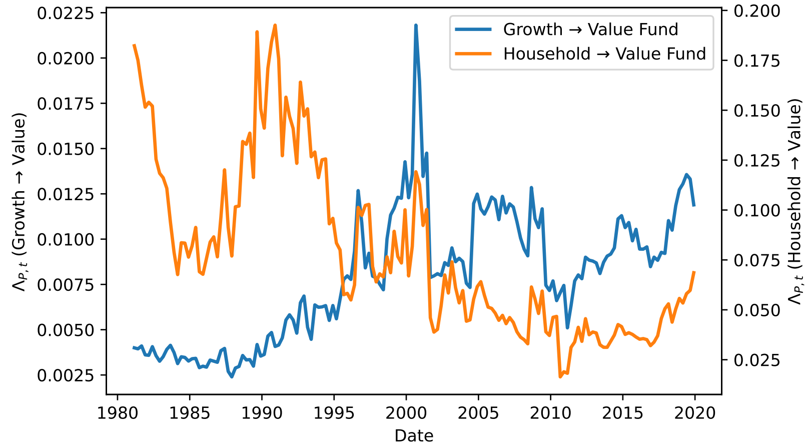
Panel (a): Value capacity in dollars ($C^*_{P,t}$)



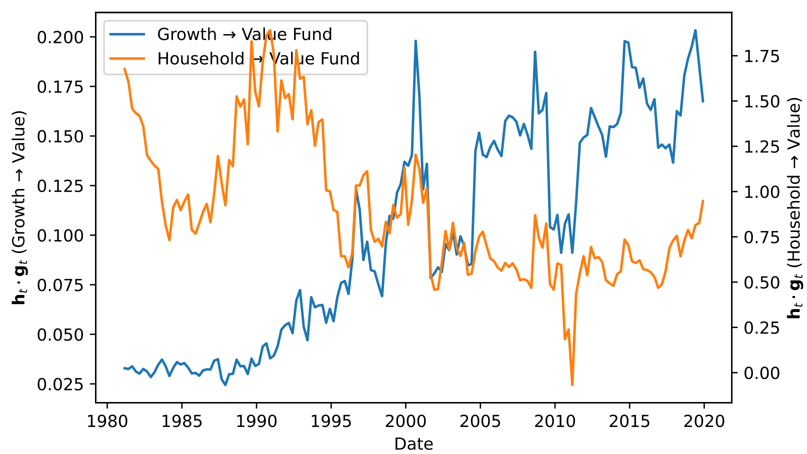
Panel (b): Value capacity as a fraction of donor AUM ($C^*_{P,t}/A_{D,t}$)

This figure plots the time series of *value capacity* under the counterfactual that reallocates capital from growth funds or households (donor group D) to value funds (recipient group R). Capacity $C^*_{P,t}$ is defined as the dollar reallocation required to drive the value strategy's alpha to zero at time t. Panel A reports $C^*_{P,t}$ in dollars. Panel B reports the same object scaled by donor AUM, $C^*_{P,t}/A_{D,t}$. The sample covers 1981–2019.

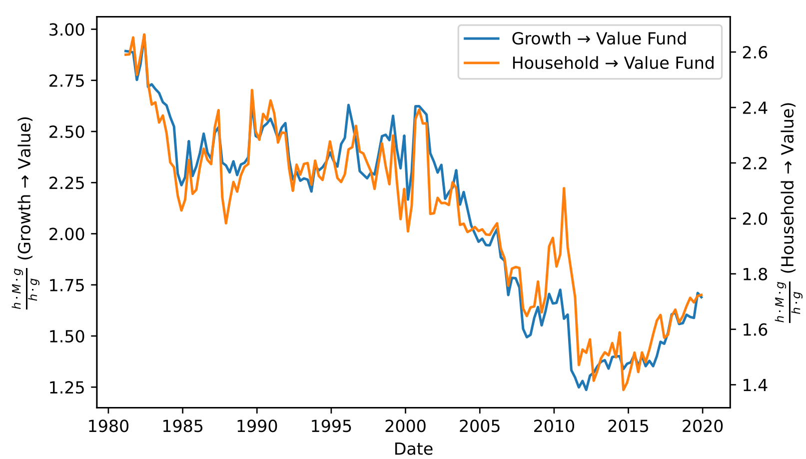
Figure 5: Decomposing time variation in alpha sensitivity



Panel (a): Alpha sensitivity $\Lambda_{P,t}$



Panel (b): Raw alignment between strategy and reallocation, $\mathbf{h}_t'\mathbf{g}_t$



Panel (c): Multiplier amplification, $\dfrac{\mathbf{h}_t'\mathbf{M}_t\mathbf{g}_t}{\mathbf{h}_t'\mathbf{g}_t}$

This figure decomposes the time variation in the alpha sensitivity $\Lambda_{P,t}$ that governs alpha rewriting in Proposition 3, $\Delta\alpha_{P,t} \approx -(C/G_t)\Lambda_{P,t}$ with $\Lambda_{P,t} = \mathbf{h}_t'\mathbf{D}(\boldsymbol{\theta})\mathbf{M}_t\mathbf{g}_t$. Panel A plots $\Lambda_{P,t}$ over time for two reallocation scenarios (growth→value versus households→value; same dollar amount). Panel B isolates the "raw" channel by plotting $\mathbf{h}_t'\mathbf{g}_t$, the alignment between the strategy alpha weights $\mathbf{h}_t$ and the donor–recipient trade list $\mathbf{g}_t$. Panel C isolates the equilibrium amplification from the demand system by plotting the ratio $(\mathbf{h}_t'\mathbf{M}_t\mathbf{g}_t)/(\mathbf{h}_t'\mathbf{g}_t)$: values above one indicate that equilibrium multipliers magnify the effect of a given trade list through inelastic demand and cross-asset spillovers. All series are computed at the quarterly frequency.

Table 1: Aggregate demand elasticities and price multipliers: own vs. cross components

This table summarizes the estimated *aggregate* elasticity matrix $\mathbf{E}_t^{\text{agg}} = \sum_i \mathbf{S}_{i,t}\widehat{\mathbf{E}}_{i,t}$ and its inverse, the multiplier matrix $\mathbf{M}_t = (\mathbf{E}_t^{\text{agg}})^{-1}$ (Section 3.2). For each quarter t, we compute cross-sectional percentiles of the diagonal (own) elements and the off-diagonal (cross) elements. The reported numbers are time-series averages of these within-quarter percentiles across all quarters in the sample.

	p5	p25	median	mean	p75	p95
Panel A: Aggregate elasticities $\mathbf{E}_t^{\text{agg}}$						
Own elasticity diag($\mathbf{E}_t^{\text{agg}}$)	0.381	0.409	0.499	0.552	0.686	0.845
Cross elasticity offdiag($\mathbf{E}_t^{\text{agg}}$)	-8.56×10^{-6}	-4.33×10^{-6}	-2.04×10^{-6}	-2.88×10^{-6}	-6.21×10^{-7}	-3.18×10^{-8}
Panel B: Multipliers $\mathbf{M}_t = (\mathbf{E}_t^{\text{agg}})^{-1}$						
Own multiplier diag($\mathbf{M}_t$)	1.663	1.957	2.347	2.339	2.737	2.968
Cross multiplier offdiag($\mathbf{M}_t$)	1.46×10^{-7}	2.80×10^{-6}	9.10×10^{-6}	1.35×10^{-5}	1.98×10^{-5}	4.19×10^{-5}

Table 2: 2SLS estimates of pass-through θ using price-wedge instruments

This table reports 2SLS estimates of the pass-through parameter θ from Section 3.3. The second-stage dependent variable is the one-period-ahead log total return (including dividends), $r_{i,t}$. Endogenous regressors are the lagged log price change $\Delta p_{i,t-1}$ and the lagged valuation ratio $v_{i,t-1} \equiv p_{i,t-1} - b_{i,t-1}$. Instruments are $z^{\Delta}_{i,t-1} \equiv \Delta PW_{i,t-1}$ and $z^{L}_{i,t-1} \equiv PW_{i,t-1}$. Columns report estimates for (i) all stocks, (ii) the growth (L) and value (H) book-to-market deciles, and (iii) the loser (L) and winner (H) momentum deciles. Pass-through is computed as $\widehat{\theta} = -\big(\widehat{\beta}_{\Delta p} + \widehat{\beta}_{v}\big)$. t-statistics (in brackets) are based on standard errors clustered by firm and time.

	All stocks	Value (B/M)		Momentum	
		L (Growth)	H (Value)	L (Losers)	H (Winners)
Panel A: First stages					
$\Delta p_{i,t-1}$ on $z^{\Delta}_{i,t-1}$	3.095	1.747	4.108	0.278	3.938
	[25.693]	[15.736]	[14.892]	[12.320]	[14.823]
$\Delta p_{i,t-1}$ on $z^{L}_{i,t-1}$	0.126	0.618	0.775	-0.037	0.234
	[2.732]	[7.084]	[8.867]	[-1.919]	[1.232]
$v_{i,t-1}$ on $z^{\Delta}_{i,t-1}$	1.315	1.998	2.246	0.331	0.363
	[36.394]	[23.893]	[26.010]	[3.590]	[4.425]
$v_{i,t-1}$ on $z^{L}_{i,t-1}$	3.526	0.237	1.390	3.702	4.003
	[75.986]	[2.429]	[16.547]	[62.036]	[48.457]
KP Wald F	339.384	11.164	54.153	75.263	108.268
Panel B: Second stage (2SLS), dependent variable $r_{i,t}$					
$\widehat{\Delta p}_{i,t-1}$	0.005	-0.113	0.049	-0.383	-0.054
	[0.353]	[-0.754]	[0.881]	[-1.184]	[-3.558]
$\widehat{v}_{i,t-1}$	-0.047	0.060	-0.113	-0.075	-0.046
	[-4.381]	[0.431]	[-1.376]	[-3.637]	[-2.674]
Pass-through $\widehat{\theta} = -\big(\widehat{\beta}_{\Delta p} + \widehat{\beta}_{v}\big)$	**0.042**	**0.052**	**0.064**	**0.458**	**0.100**
	[3.332]	**[1.974]**	**[1.969]**	**[1.467]**	**[5.407]**
Observations	528,916	53,413	51,547	50,502	52,640

Table 3: Summary statistics for value and momentum decile portfolios

This table reports annualized average excess returns $\overline{R}$, CAPM alphas α^{CAPM} (in %), and market betas β^{MKT} for decile portfolios formed over 1981–2019. Panel A sorts stocks on book-to-market (B/M) at the end of June (L = Growth, H = Value). Panel B sorts stocks on momentum measured as cumulative returns from month $t-12$ to $t-2$ (L = Losers, H = Winners). H–L is the corresponding long–short strategy. Brackets report t-statistics based on White (1980) heteroskedasticity-consistent standard errors.

	L	2	3	4	5	6	7	8	9	H	**H–L**
Panel A: Book-to-market (Value)											
$\overline{R}$	7.704	6.960	8.424	7.524	8.604	8.736	8.904	8.220	9.552	12.516	**4.812**
	[2.63]	[2.7]	[3.4]	[2.96]	[3.4]	[3.57]	[3.65]	[3.22]	[3.81]	[4.04]	**[1.81]**
α^{CAPM}	-1.020	-0.948	0.780	-0.240	1.044	1.488	1.956	1.176	2.676	4.572	**5.592**
	[-0.86]	[-1.1]	[0.99]	[-0.28]	[1.02]	[1.4]	[1.54]	[0.8]	[1.85]	[2.28]	**[2.0]**
β^{MKT}	1.113	1.009	0.976	0.991	0.965	0.924	0.886	0.899	0.876	1.013	**-0.100**
	[42.35]	[57.81]	[50.05]	[44.9]	[41.2]	[33.24]	[27.37]	[22.54]	[25.52]	[20.94]	[-1.52]
Panel B: Momentum											
$\overline{R}$	-0.996	4.620	6.396	8.376	7.596	6.936	8.220	8.436	8.904	12.624	**13.608**
	[-0.19]	[1.22]	[1.98]	[3.08]	[3.05]	[2.94]	[3.57]	[3.51]	[3.36]	[3.43]	**[2.99]**
α^{CAPM}	-13.788	-5.184	-2.388	0.624	0.300	-0.108	1.332	1.272	1.284	3.108	**16.896**
	[-4.36]	[-2.25]	[-1.32]	[0.47]	[0.28]	[-0.12]	[1.47]	[1.35]	[1.0]	[1.39]	**[3.92]**
β^{MKT}	1.633	1.251	1.120	0.988	0.931	0.899	0.879	0.914	0.973	1.213	**-0.420**
	[19.74]	[20.23]	[20.47]	[26.2]	[36.02]	[42.18]	[35.75]	[46.08]	[29.55]	[24.84]	[-3.79]

Table 4: How reallocating growth-fund capital rewrites value and momentum alphas

This table reports the impact of the *same* dollar reallocation on two long–short anomalies: value (H–L sorted on book-to-market) and momentum (H–L sorted on past returns). In both scenarios, the reallocation amount is fixed at 100% of aggregate growth-fund AUM and is transferred into value funds. The source of the outflow differs: (i) growth funds themselves, or (ii) households. The first two columns report the implied change in the value strategy's annual expected return and CAPM alpha; the last two columns report the implied change in the momentum strategy under the same reallocations (a cross-anomaly spillover). Brackets report t-statistics. Capacity C^* is reported only for value and is defined as the dollar reallocation required to drive the value alpha to zero.

	Value (B/M) strategy		**Momentum strategy**	
	$\mathbb{E}[r^{H-L}]$	α^{CAPM}	$\mathbb{E}[r^{H-L}]$	α^{CAPM}
Baseline (no reallocation)				
	4.812	5.592	13.608	16.896
Reallocation: $C = 100\%$ **of growth-fund AUM is transferred**				
Growth funds → Value funds	-0.735	-0.736	0.414	0.410
	[-8.96]	[-8.95]	[6.26]	[6.32]
Households → Value funds	-0.411	-0.418	0.090	0.062
	[-6.78]	[-6.84]	[3.17]	[2.24]
Post-reallocation (baseline + change)				
Growth funds → Value funds	4.077	4.856	14.022	17.306
Households → Value funds	4.401	5.174	13.698	16.958
Implied value capacity C^* **(dollars required to set** $\alpha^{CAPM}_{value} = 0$**)**				
C^* (billion USD), Growth funds → Value funds	3055.401	(896.159% of growth-fund AUM)		
C^* (billion USD), Households → Value funds	5005.506	(94.432% of household equity AUM)		

Table 5: Distribution of holdings across book-to-market deciles

This table reports, by decade subsample, how value/growth mutual funds and households allocate across book-to-market (B/M) deciles. For each investor type and period, *Fraction of Dollar Value* is the share of the investors total dollar holdings invested in each B/M decile. *Percentage of Stocks* is the share of stocks in each B/M decile held by the investors (equal-weighted across stocks within each decile).

		L	2	3	4	5	6	7	8	9	H
Panel A: Value funds											
Fraction of dollar value	1981–1990	0.14%	0.24%	0.38%	0.46%	0.55%	0.65%	0.81%	1.02%	1.27%	1.31%
Percentage of stocks		14.99%	29.04%	34.54%	39.75%	39.65%	41.78%	44.06%	43.89%	41.81%	31.88%
Fraction of dollar value	1991–2000	0.75%	1.73%	2.28%	2.76%	3.23%	3.31%	3.57%	3.74%	4.06%	4.21%
Percentage of stocks		36.58%	51.27%	61.65%	66.49%	69.87%	71.65%	73.03%	73.01%	69.84%	62.37%
Fraction of dollar value	2001–2010	1.92%	3.21%	4.18%	4.88%	5.25%	5.76%	6.16%	6.30%	6.39%	6.60%
Percentage of stocks		71.11%	79.18%	83.78%	86.61%	85.68%	85.81%	85.38%	84.07%	84.67%	84.63%
Fraction of dollar value	2011–2019	1.34%	2.01%	2.63%	3.42%	3.98%	4.50%	5.02%	5.35%	5.58%	5.72%
Percentage of stocks		71.75%	83.16%	87.68%	89.30%	89.35%	90.57%	91.34%	89.25%	85.43%	79.34%
Panel B: Growth funds											
Fraction of dollar value	1981–1990	2.15%	1.32%	1.03%	0.89%	0.77%	0.60%	0.47%	0.36%	0.34%	0.42%
Percentage of stocks		42.74%	55.93%	52.37%	46.86%	41.51%	37.44%	32.44%	28.35%	24.01%	16.47%
Fraction of dollar value	1991–2000	3.85%	2.84%	2.45%	2.14%	1.58%	1.47%	1.28%	1.05%	0.89%	1.00%
Percentage of stocks		60.22%	69.03%	69.12%	64.70%	60.03%	52.35%	44.93%	39.69%	31.73%	20.00%
Fraction of dollar value	2001–2010	7.81%	6.28%	4.85%	4.32%	3.65%	3.45%	3.23%	3.08%	2.95%	3.01%
Percentage of stocks		84.30%	86.66%	85.44%	83.67%	77.93%	73.13%	67.15%	59.50%	54.54%	45.56%
Fraction of dollar value	2011–2019	8.67%	6.67%	5.21%	4.57%	3.66%	2.81%	2.34%	1.89%	1.63%	1.53%
Percentage of stocks		85.27%	88.94%	88.65%	86.61%	84.04%	82.69%	80.91%	74.90%	62.47%	45.36%
Panel C: Households											
Fraction of Dollar Value	1981-03 – 1990-12	56.24%	50.62%	50.1%	52.09%	53.49%	55.6%	56.97%	58.26%	59.24%	58.57%
Percentage of Stocks		99.97%	99.97%	99.94%	99.98%	99.89%	99.93%	99.98%	99.98%	99.96%	99.94%
Fraction of Dollar Value	1991-03 – 2000-12	47.2%	44.16%	41.93%	42.13%	44.17%	45.49%	46.67%	47.16%	45.81%	48.32%
Percentage of Stocks		99.17%	99.3%	99.43%	99.47%	99.62%	99.66%	99.72%	99.74%	99.86%	99.76%
Fraction of Dollar Value	2001-03 – 2010-12	30.18%	29.12%	30.15%	28.71%	31.47%	29.13%	29.43%	30.69%	29.49%	29.09%
Percentage of Stocks		95.27%	94.99%	95.2%	95.27%	95.89%	95.98%	96.26%	96.77%	96.97%	97.46%
Fraction of Dollar Value	2011-03 – 2019-12	18.16%	17.45%	21.67%	18.1%	19.2%	19.21%	18.98%	20.71%	20.02%	21.64%
Percentage of Stocks		97.44%	97.6%	97.88%	98.45%	98.49%	99.06%	99.09%	99.21%	98.89%	98.77%

Online Appendix

A Data Description

We obtain monthly stock prices and returns from the Center for Research in Security Prices (CRSP) for all common stocks (share codes 10 and 11) listed on NYSE, AMEX, and NASDAQ between January 1981 and December 2019. Firm-level accounting variables, including book equity, come from Compustat. For mutual funds, we rely on the CRSP Survivor-Bias-Free Mutual Fund Database for characteristics such as expense ratios and turnover.

A.1 Mutual Fund Selection and Classification.

Following Kacperczyk et al. (2008) and Evans (2010), we first screen the CRSP mutual fund data to identify domestic equity mutual funds, merging multiple share classes (CRSP_FUNDNO) into single-fund entities (WFICN) using the MFLINKS file from Wharton Research Data Services. We exclude funds that, on average, invest less than 80% of their assets in equity, as well as those whose total net assets remain below $15 million. To isolate actively managed equity funds, we remove index funds following the procedure of Dannhauser and Pontiff (2024), which identifies index-tracking portfolios by searching the CRSP fund names for key strings.[11]

We then merge these active funds with their Morningstar Direct style categories, narrowing our focus to nine widely studied Morningstar styles defined by size (small, mid, large) and style tilt (value, blend, growth). This classification scheme mirrors that in Han et al. (2021) and ensures consistency in style-based holdings analysis.

A.2 Classification of Other Investors

In addition to these nine Morningstar-based active mutual fund groups, we classify all remaining equity investors into one of several categories: index mutual funds, banks, insurance

[11] An index fund is identified if its CRSP fund name contains "SP," "DOW," "Dow," "DJ," or if the name (in lowercase) contains any of the following: "index," "idx," "indx," "ind_," "composite," "russell," "s&p," "s and p," "s & p," "msci," "Bloomberg," "kbw," "nasdaq," "nyse," "stoxx," "ftse," "wilshire," "Morningstar," or numeric labels such as "100," "400," "500," "600," "900," "1000," "1500," "2000," "3000," or "5000."

companies, investment advisors, pension funds, university endowments and foundations, or households. Thomson Reuters Institutional Holdings Database (s34 file) assigns each manager to a type code (Variable Name TYPECODE),

1. Bank,
2. Insurance companies,
3. Investment company and their managers,
4. Investment advisors, and
5. All other (i.e., pension funds, university endowments, and foundations).

Unfortunately, the TYPECODE variable is not reliable from 1998 and beyond, when many institutions were and are improperly classified as type 5 (endowments and others).[12] We correct the type codes through the following steps:

1. If a managers' first type code is not 5, we replace the incorrect type code of 5 in the subsequent periods with the first type code.
2. Following KY, we assign type codes 3 through 5 to pension funds using the list of top 300 pension funds (Watson (2015)).
3. As in KY, we construct a database of investment advisors based on the historical archives of Securities and Exchange Commission Form ADV since June 2006. We use the Fuzzy Matching function in Python to match manager names to primary business names in the investment advisor database. If a valid match is found, we reclassify type code 5 to type 4.

A.3 Holdings Data.

We obtain the mutual fund holdings data from Thomson Reuters mutual fund holdings (s12 files), and use Thomson Reuters institutional holdings (s34 files) to capture other institutional investors. Note that the s34 files aggregate all mutual fund holdings at the family

[12] (See the user's guide provided by Wharton Research Data Services in 2008)

level, so to avoid double counting, we merge the s12 type5 files with s34 records, and exclude any duplicated mutual fund positions. We then follow KY by defining household holdings for a given stock as the remainder of shares outstanding once we subtract the combined holdings of all mutual funds and other institutional investors. Under this definition, smaller institutions (with AUM below \$100 million and thus not required to file with the SEC) are commingled with households, as they are unobserved in the institutional holdings files.

B Proofs and Other Derivations

B.1 Deriving the Aggregate Elasticity Matrix from First-Order Market Clearing

This subsection derives Equation (6) in the main text. Start from exact market clearing in logs at time t:

$$\mathbf{H}(\mathbf{p}_t, \mathbf{a}_t) \equiv \log(\mathbf{Q}_t^s) - \log(\mathbf{Q}_t^d(\mathbf{p}_t, \mathbf{a}_t)) = \mathbf{0},$$

where, stock-by-stock,

$$Q^d_{n,t}(\mathbf{p}_t, \mathbf{a}_t) = \sum_i \exp\Big(a_{i,t} + k_{n,i,t} - (\mathbf{E}_{i,t}\mathbf{p}_t)_n\Big),$$

and $\mathbf{k}_{i,t}$ is treated as fixed for this counterfactual.

Now take one first-order Taylor expansion of $\mathbf{H}$ around the observed equilibrium $(\mathbf{p}_t, \mathbf{a}_t)$, with perturbations $\Delta\mathbf{p}_t$ and $\Delta\mathbf{a}_t$:

$$\mathbf{0} = \mathbf{H}(\mathbf{p}_t + \Delta\mathbf{p}_t, \mathbf{a}_t + \Delta\mathbf{a}_t) \approx \mathbf{H}(\mathbf{p}_t, \mathbf{a}_t) + \mathbf{H}_p\,\Delta\mathbf{p}_t + \mathbf{H}_a\,\Delta\mathbf{a}_t.$$

Because $\mathbf{H}(\mathbf{p}_t, \mathbf{a}_t) = \mathbf{0}$, this becomes

$$\mathbf{H}_p\,\Delta\mathbf{p}_t + \mathbf{H}_a\,\Delta\mathbf{a}_t \approx \mathbf{0}.$$

Compute these Jacobians at the observed equilibrium. Define ownership shares

$$s_{n,i,t} \equiv \frac{Q^d_{n,i,t}}{Q^d_{n,t}} = \frac{Q^d_{n,i,t}}{Q^s_{n,t}}, \qquad \mathbf{S}_{i,t} \equiv \operatorname{diag}(s_{1,i,t}, \ldots, s_{N,i,t}).$$

Then

$$\frac{\partial \log Q^d_{n,t}}{\partial p_{m,t}} = -\sum_i s_{n,i,t}(\mathbf{E}_{i,t})_{n,m},$$

so

$$\mathbf{H}_p = \sum_i \mathbf{S}_{i,t}\mathbf{E}_{i,t} \equiv \mathbf{E}_t^{\text{agg}}.$$

Also,

$$\frac{\partial \log Q^d_{n,t}}{\partial a_{i,t}} = s_{n,i,t},$$

hence $(\mathbf{H}_a\Delta\mathbf{a}_t)_n = -\sum_i s_{n,i,t}\Delta a_{i,t}$, i.e.

$$\mathbf{H}_a\Delta\mathbf{a}_t = -\sum_i \mathbf{S}_{i,t}\Delta a_{i,t}\mathbf{1} \equiv -\mathbf{f}_t.$$

Substitute into the linearized system:

$$\mathbf{E}_t^{\text{agg}}\Delta\mathbf{p}_t - \mathbf{f}_t \approx \mathbf{0} \quad \Rightarrow \quad \mathbf{E}_t^{\text{agg}}\Delta\mathbf{p}_t \approx \mathbf{f}_t.$$

If $\mathbf{E}_t^{\text{agg}}$ is invertible,

$$\Delta\mathbf{p}_t \approx (\mathbf{E}_t^{\text{agg}})^{-1}\mathbf{f}_t \equiv \mathbf{M}_t\mathbf{f}_t,$$

which is Equation (9), with

$$\mathbf{E}_t^{\text{agg}} = \sum_i \mathbf{S}_{i,t}\mathbf{E}_{i,t},$$

as in Equation (6).

B.2 Donor–Recipient Ownership Overlap and Linearity in Reallocation Size

This subsection derives Equation (14) in the main text. Suppose C dollars are reallocated from donor group D to recipient group R, with proportional scaling within each group. Let

$$A_{D,t} \equiv \sum_{i\in D} A_{i,t}, \qquad A_{R,t} \equiv \sum_{i\in R} A_{i,t}.$$

Then dollar AUM changes are

$$\Delta A_{i,t} = \begin{cases} -C \cdot \dfrac{A_{i,t}}{A_{D,t}}, & i \in D, \\ \quad C \cdot \dfrac{A_{i,t}}{A_{R,t}}, & i \in R, \\ 0, & \text{otherwise.} \end{cases}$$

For small reallocations, $\Delta a_{i,t} = \Delta \log A_{i,t} \approx \Delta A_{i,t}/A_{i,t}$, so all investors within the same group share the same log-AUM shock:

$$\Delta a_{i,t} \approx \begin{cases} -\dfrac{C}{A_{D,t}}, & i \in D, \\ \quad \dfrac{C}{A_{R,t}}, & i \in R, \\ 0, & \text{otherwise.} \end{cases}$$

Using the main-text definition

$$\mathbf{f}_t = \sum_i \mathbf{S}_{i,t} \Delta a_{i,t} \mathbf{1},$$

we obtain

$$\mathbf{f}_t \approx C \left(-\frac{1}{A_{D,t}} \sum_{i \in D} \mathbf{S}_{i,t} \mathbf{1} + \frac{1}{A_{R,t}} \sum_{i \in R} \mathbf{S}_{i,t} \mathbf{1} \right). \tag{B.1}$$

Hence $\mathbf{f}_t$ is linear in C.

Now choose any positive normalization G_t, and define

$$\mathbf{g}_t \equiv G_t \left(-\frac{1}{A_{D,t}} \sum_{i \in D} \mathbf{S}_{i,t} \mathbf{1} + \frac{1}{A_{R,t}} \sum_{i \in R} \mathbf{S}_{i,t} \mathbf{1} \right).$$

Then (B.1) is exactly

$$\mathbf{f}_t \approx \frac{C}{G_t} \mathbf{g}_t,$$

which is Equation (14) in the main text.

For stock n, use

$$s_{n,i,t} = \frac{Q^d_{n,i,t}}{Q^s_{n,t}} = \frac{w_{n,i,t}A_{i,t}}{P_{n,t}Q^s_{n,t}} = \frac{w_{n,i,t}A_{i,t}}{\mathrm{ME}_{n,t}}, \quad \mathrm{ME}_{n,t} \equiv P_{n,t}Q^s_{n,t},$$

to write

$$g_{n,t} = \frac{G_t}{\mathrm{ME}_{n,t}} \left(\bar{w}_{n,R,t} - \bar{w}_{n,D,t}\right),$$

where

$$\bar{w}_{n,R,t} \equiv \sum_{i \in R} \left(\frac{A_{i,t}}{A_{R,t}}\right) w_{n,i,t}, \qquad \bar{w}_{n,D,t} \equiv \sum_{i \in D} \left(\frac{A_{i,t}}{A_{D,t}}\right) w_{n,i,t}.$$

So $\mathbf{g}_t$ is the donor–recipient difference in group-level portfolio weights, scaled by inverse market equity.

B.3 Proof of Proposition 1

Proof. From the first-order market-clearing map in Appendix B.1, there exists a constant vector $\mathbf{b}_t$ such that

$$\mathbf{p}_t = \mathbf{M}_t \left(\mathbf{b}_t + \sum_i \mathbf{S}_{i,t}(a_{i,t}\mathbf{1} + \mathbf{k}_{i,t})\right), \qquad \mathbf{M}_t \equiv (\mathbf{E}_t^{\mathrm{agg}})^{-1}.$$

The counterfactual equilibrium satisfies the analogous expression:

$$\mathbf{p}_t^{CF} = \mathbf{M}_t^{CF} \left(\mathbf{b}_t + \sum_i \mathbf{S}_{i,t}^{CF}(a_{i,t}^{CF}\mathbf{1} + \mathbf{k}_{i,t})\right),$$

where $\mathbf{k}_{i,t}$ is held fixed by construction of the counterfactual.

Subtracting gives

$$\Delta\mathbf{p}_t = \mathbf{M}_t \left[\sum_i \mathbf{S}_{i,t}^{CF}(a_{i,t}^{CF}\mathbf{1} + \mathbf{k}_{i,t}) - \sum_i \mathbf{S}_{i,t}(a_{i,t}\mathbf{1} + \mathbf{k}_{i,t})\right] + \Delta\mathbf{M}_t \left[\mathbf{b}_t + \sum_i \mathbf{S}_{i,t}^{CF}(a_{i,t}^{CF}\mathbf{1} + \mathbf{k}_{i,t})\right].$$

Using $\mathbf{S}_{i,t}^{CF} = \mathbf{S}_{i,t} + \Delta\mathbf{S}_{i,t}$ and $a_{i,t}^{CF} = a_{i,t} + \Delta a_{i,t}$,

$$\begin{aligned}\sum_i \mathbf{S}_{i,t}^{CF}(a_{i,t}^{CF}\mathbf{1} + \mathbf{k}_{i,t}) - \sum_i \mathbf{S}_{i,t}(a_{i,t}\mathbf{1} + \mathbf{k}_{i,t}) = & \sum_i \mathbf{S}_{i,t}\,\Delta a_{i,t}\mathbf{1} \\ & + \sum_i \Delta\mathbf{S}_{i,t}\,(a_{i,t}\mathbf{1} + \mathbf{k}_{i,t}) \\ & + \sum_i \Delta\mathbf{S}_{i,t}\,\Delta a_{i,t}\mathbf{1}.\end{aligned}$$

The last line is second order ($\Delta\mathbf{S} \cdot \Delta a$) and is dropped under the first-order approximation. Replacing $\mathbf{b}_t + \sum_i \mathbf{S}_{i,t}^{CF}(a_{i,t}^{CF}\mathbf{1} + \mathbf{k}_{i,t})$ in the $\Delta\mathbf{M}_t$ term by its baseline counterpart $\sum_i \mathbf{S}_{i,t}(a_{i,t}\mathbf{1} + \mathbf{k}_{i,t})$ is again first-order accurate because $\Delta\mathbf{M}_t$ already multiplies the bracket.

Therefore

$$\Delta\mathbf{p}_t = \underbrace{\mathbf{M}_t\left(\sum_i \mathbf{S}_{i,t}\,\Delta a_{i,t}\mathbf{1}\right)}_{\text{direct effect}} + \underbrace{\mathbf{M}_t\left(\sum_i \Delta\mathbf{S}_{i,t}\,(a_{i,t}\mathbf{1} + \mathbf{k}_{i,t})\right)}_{\text{share re-composition}} + \underbrace{\Delta\mathbf{M}_t\left(\sum_i \mathbf{S}_{i,t}\,(a_{i,t}\mathbf{1} + \mathbf{k}_{i,t})\right)}_{\text{multiplier change}},$$

which is Equation (8). Defining $\mathbf{f}_t \equiv \sum_i \mathbf{S}_{i,t}\Delta a_{i,t}\mathbf{1}$, if the last two terms are quantitatively negligible, we obtain

$$\Delta\mathbf{p}_t \approx \mathbf{M}_t\mathbf{f}_t,$$

which is Equation (9). ■

B.4 Proof of Proposition 2

Proof. Start from the Campbell–Shiller identity in Equation (10):

$$r_{n,t+1} = \kappa_n + (1-\rho_n)d_{n,t+1} + \rho_n p_{n,t+1} - p_{n,t}.$$

Taking conditional expectations gives

$$\mu_{n,t} = \kappa_n + (1-\rho_n)\mathbb{E}_t[d_{n,t+1}] + \rho_n\mathbb{E}_t[p_{n,t+1}] - p_{n,t}.$$

Differentiate with respect to the current log price $p_{n,t}$:

$$\frac{\partial \mu_{n,t}}{\partial p_{n,t}} = (1-\rho_n)\frac{\partial \mathbb{E}_t[d_{n,t+1}]}{\partial p_{n,t}} + \rho_n \frac{\partial \mathbb{E}_t[p_{n,t+1}]}{\partial p_{n,t}} - 1.$$

Under the proposition's identifying restriction $\frac{\partial \mathbb{E}_t[d_{n,t+1}]}{\partial p_{n,t}} = 0$ and with $\psi_n \equiv \frac{\partial \mathbb{E}_t[p_{n,t+1}]}{\partial p_{n,t}}$ from Equation (11),

$$\frac{\partial \mu_{n,t}}{\partial p_{n,t}} = \rho_n \psi_n - 1 = -(1-\rho_n\psi_n) = -\theta_n.$$

Hence, for a small equilibrium price change $\Delta p_{n,t}$,

$$\Delta \mu_{n,t} \approx -\theta_n \, \Delta p_{n,t},$$

which is Equation (12).

Stacking across assets,

$$\Delta \boldsymbol{\mu}_t \approx -\mathbf{D}(\boldsymbol{\theta})\, \Delta \mathbf{p}_t.$$

Using Proposition 1,

$$\Delta \mathbf{p}_t \approx \mathbf{M}_t \mathbf{f}_t,$$

so

$$\Delta \boldsymbol{\mu}_t \approx -\mathbf{D}(\boldsymbol{\theta})\, \mathbf{M}_t\, \mathbf{f}_t,$$

which is Equation (13). ■

B.5 Proof of Proposition 3

Proof. Let

$$\alpha_{P,t} = \mathbf{h}_t' \boldsymbol{\mu}_t, \qquad \mathbf{h}_t = \mathbf{w}_{L,t} - \mathbf{w}_{S,t} - (\beta_L - \beta_S)\mathbf{w}_{M,t},$$

as defined in Equation (15). Holding $\mathbf{h}_t$ fixed within the counterfactual,

$$\Delta \alpha_{P,t} = \mathbf{h}_t' \Delta \boldsymbol{\mu}_t.$$

From Proposition 2,

$$\Delta \boldsymbol{\mu}_t \approx -\mathbf{D}(\boldsymbol{\theta})\, \mathbf{M}_t\, \mathbf{f}_t.$$

From Equation (14), for a $D \to R$ reallocation of C dollars,

$$\mathbf{f}_t \approx \frac{C}{G_t}\mathbf{g}_t.$$

Substitute both expressions:

$$\Delta\alpha_{P,t} \approx \mathbf{h}_t' \left(-\mathbf{D}(\boldsymbol{\theta})\,\mathbf{M}_t\,\frac{C}{G_t}\mathbf{g}_t \right) = -\frac{C}{G_t}\mathbf{h}_t'\mathbf{D}(\boldsymbol{\theta})\mathbf{M}_t\mathbf{g}_t.$$

Define

$$\Lambda_{P,t} \equiv \mathbf{h}_t'\mathbf{D}(\boldsymbol{\theta})\mathbf{M}_t\mathbf{g}_t,$$

to obtain

$$\Delta\alpha_{P,t} \approx -\frac{C}{G_t}\Lambda_{P,t},$$

which is Equation (16). ■

B.6 Detailed Capacity Computation

This subsection provides a full derivation of the capacity formula in the main text. The goal is to show, step by step, how a dollar reallocation C maps into $\Delta\alpha_{P,t}$, and then how we solve for the C that sets alpha to zero.

Step 1: Reallocation rule in investor log AUM. Move C dollars from donor group D to recipient group R, with proportional scaling within each group:

$$\Delta A_{i,t} = \begin{cases} -C \cdot \dfrac{A_{i,t}}{A_{D,t}}, & i \in D, \\ C \cdot \dfrac{A_{i,t}}{A_{R,t}}, & i \in R, \\ 0, & \text{otherwise.} \end{cases}$$

For small reallocations, $\Delta a_{i,t} = \Delta \log A_{i,t} \approx \Delta A_{i,t}/A_{i,t}$, so

$$\Delta a_{i,t} \approx \begin{cases} -\dfrac{C}{A_{D,t}}, & i \in D, \\ \dfrac{C}{A_{R,t}}, & i \in R, \\ 0, & \text{otherwise.} \end{cases}$$

Step 2: Ownership-weighted demand shock is linear in C. Using $\mathbf{f}_t \equiv \sum_i \mathbf{S}_{i,t} \Delta a_{i,t} \mathbf{1}$,

$$\mathbf{f}_t \approx C \left(-\frac{1}{A_{D,t}} \sum_{i \in D} \mathbf{S}_{i,t} \mathbf{1} + \frac{1}{A_{R,t}} \sum_{i \in R} \mathbf{S}_{i,t} \mathbf{1} \right) \equiv \frac{C}{G_t} \mathbf{g}_t,$$

where $\mathbf{g}_t$ is defined in Equation (14). So the cross-sectional trade direction is $\mathbf{g}_t$, and C/G_t sets its scale.

Step 3: Price impact map. From Proposition 1,

$$\Delta \mathbf{p}_t \approx \mathbf{M}_t \mathbf{f}_t \approx \frac{C}{G_t} \mathbf{M}_t \mathbf{g}_t.$$

Step 4: Expected-return rewriting. From Proposition 2,

$$\Delta \boldsymbol{\mu}_t \approx -\mathbf{D}(\boldsymbol{\theta})\, \Delta \mathbf{p}_t \approx -\frac{C}{G_t}\, \mathbf{D}(\boldsymbol{\theta})\, \mathbf{M}_t\, \mathbf{g}_t.$$

Step 5: Strategy alpha rewriting. By definition of alpha weights $\mathbf{h}_t$ in Equation (15),

$$\Delta \alpha_{P,t} = \mathbf{h}_t' \Delta \boldsymbol{\mu}_t \approx -\frac{C}{G_t}\, \mathbf{h}_t' \mathbf{D}(\boldsymbol{\theta}) \mathbf{M}_t \mathbf{g}_t.$$

Define

$$\Lambda_{P,t} \equiv \mathbf{h}_t' \mathbf{D}(\boldsymbol{\theta}) \mathbf{M}_t \mathbf{g}_t,$$

to obtain

$$\Delta \alpha_{P,t} \approx -\frac{C}{G_t}\, \Lambda_{P,t}, \tag{B.2}$$

which is Equation (16).

Step 6: Solve for capacity. Capacity $C^*_{P,t}$ is the reallocation that sets counterfactual alpha to zero:

$$\bar{\alpha}_{P,t} + \Delta\alpha_{P,t} = 0.$$

Substitute (B.2):

$$\bar{\alpha}_{P,t} - \frac{C^*_{P,t}}{G_t}\Lambda_{P,t} = 0,$$

hence

$$C^*_{P,t} = \left(\frac{G_t}{\Lambda_{P,t}}\right)\bar{\alpha}_{P,t}. \tag{B.3}$$

This is exactly Equation (17) in the main text.

If $G_t = A_{D,t}$, then $C^*_{P,t}/A_{D,t} = \bar{\alpha}_{P,t}/\Lambda_{P,t}$, so capacity is the donor outflow fraction needed to erase alpha. For an anomaly with $\bar{\alpha}_{P,t} > 0$, a finite positive capacity requires $\Lambda_{P,t} > 0$; if $\Lambda_{P,t} \leqslant 0$, this specific reallocation direction does not reduce the anomaly alpha.

C Other Details

C.1 Floating-Weight Portfolios and the Indirect Weight-Change Effect

In Section 2, we focus primarily on the direct price-change effect. Here, we show that when portfolio weights are allowed to float with prices, there is also an indirect weight-change effect. In most scenarios we analyze, however, this indirect term is quantitatively small.

C.1.1 Indirect Weight Changes in Portfolio Returns

Equation (C.1) decomposes the change in the long leg's expected return into direct and indirect components:

$$\begin{aligned}\Delta\mu_{L,t} &= \sum_n w_{n,L,t}\,\Delta\mu_{n,t} + \sum_n \Delta w_{n,L,t}\,\mu_{n,t} + \sum_n \Delta w_{n,L,t}\,\Delta\mu_{n,t} \\ &= \underbrace{-\sum_n w_{n,L,t}\,\theta^L_n\,\Delta p_{n,t}}_{\text{Direct Price-Change Effect}} \;+\; \underbrace{\sum_n \Delta w_{n,L,t}\,\mu_{n,t} \;-\; \sum_n \Delta w_{n,L,t}\,\theta^L_n\,\Delta p_{n,t}}_{\text{Indirect Weight-Change Effect}}\,.\end{aligned} \tag{C.1}$$

For a portfolio with **fixed** weights (e.g., an equal-weighted fund that actively rebalances to maintain equal shares), we have $\Delta w_{n,L,t} = 0$, so the indirect effect vanishes. But if weights are free to float as prices move, then $\Delta w_{n,L,t} \neq 0$ in general.

C.1.2 Weights Floating with Price Changes

To approximate how weights $w_{n,L,t}$ shift with price changes, assume

$$w_{n,L,t}^{CF} \;\approx\; \frac{w_{n,L,t}\,\big(1+\Delta p_{n,t}\big)}{\sum_m w_{m,L,t}\,\big(1+\Delta p_{m,t}\big)}, \tag{C.2}$$

where $\Delta p_{n,t}$ is the price change for stock n. Because $\sum_n w_{n,L,t} = 1$, we can rewrite the denominator and apply a first-order approximation for small $\Delta p_{n,t}$. Tracing through the algebra yields:

$$\Delta w_{n,L,t} \;\approx\; w_{n,L,t}\Big(\Delta p_{n,t} \;-\; \sum_m w_{m,L,t}\,\Delta p_{m,t} \;-\; \Delta p_{n,t}\sum_m w_{m,L,t}\,\Delta p_{m,t}\Big). \tag{C.3}$$

Substituting this into (C.1) yields an expanded expression for $\Delta\mu_{L,t}$, which we simplify by defining:

$$\Delta\bar{p}_{L,t} \;=\; \sum_n w_{n,L,t}\,\Delta p_{n,t}, \quad \Delta\overline{p^2_{L,t}} \;=\; \sum_n w_{n,L,t}\,\Delta p^2_{n,t}, \quad \Delta\overline{p_{L,t}\,\mu_{L,t}} \;=\; \sum_n w_{n,L,t}\,\mu_{n,t}\,\Delta p_{n,t}.$$

For constant θ_n^L across all stocks in the long leg, we obtain:

$$\Delta\mu_{L,t} \;\approx\; \underbrace{-\,\theta^L\,\Delta\bar{p}_{L,t}}_{\text{Direct Effect}} \;+\; \underbrace{\big(1-\Delta\bar{p}_{L,t}\big)\,\big(\Delta\overline{p_{L,t}\,\mu_{L,t}} - \theta^L\,\Delta\overline{p^2_{L,t}}\big) \;-\; \Delta\bar{p}_{L,t}\big(\mu_{L,t} - \theta^L\,\Delta\bar{p}_{L,t}\big)}_{\text{Weight-Adjustment Effect}}. \tag{C.4}$$

An analogous formula applies for $\Delta\mu_{S,t}$, and taking the difference gives $\Delta\mu_{P,t} = \Delta\mu_{L,t} - \Delta\mu_{S,t}$.

C.1.3 Numerical Calibration

To see how large the weight-adjustment effect can be, we run a simple calibration. From Panel A of Table 3, we use $\mu_{L,t} = 12.5\%$ and $\mu_{S,t} = 7.7\%$. We also take $\theta^L = 0.35$ and $\theta^S = 0.22$ from Panel B of the same table. Suppose we consider $(\Delta\bar{p}_{L,t}, \Delta\bar{p}_{S,t})$ pairs of

$(2.5\%, -2.5\%)$, $(5\%, -5\%)$, and $(7.5\%, -7.5\%)$. Table C.1 summarizes the result. The weight-adjustment effect contributes only a tiny fraction of the total change, reinforcing that we can typically focus on the direct effect alone.

Price Change Pair	Direct Effect	Weight-Adjustment	Combined
(0.025, -0.025)	-0.01425	-0.00002	-0.01427
(0.050, -0.050)	-0.02850	-0.00005	-0.02855
(0.075, -0.075)	-0.04275	-0.00003	-0.04278

Table C.1: Calibrated Effects for Varying Price Changes

C.1.4 One-for-One Floating

If the portfolio fully floats one-for-one with price changes (i.e., $w_{n,L,t}^{CF} \approx w_{n,L,t}(1 + \Delta p_{n,t})$, ignoring the denominator correction), the math in (C.4) simplifies further. In that case, the indirect effect becomes slightly larger, though still minor in our simulations. Table C.2 shows the calibration under this simplified assumption.

Price Change Pair	Direct Effect	Simplified WA	Combined
(0.025, -0.025)	-0.01425	0.00497	-0.00928
(0.050, -0.050)	-0.02850	0.00978	-0.01873
(0.075, -0.075)	-0.04275	0.01442	-0.02833

Table C.2: Calibrated Effects with One-for-One Floating

Overall, while weights that float with prices introduce an indirect rebalancing effect, it remains quantitatively small in our scenarios. Hence, focusing on the direct price-change channel is usually enough to capture the portfolio's counterfactual returns with high accuracy.

D Additional Results

D.1 Validation of Multiplier Estimation Using Flow Shocks

Full-Sample Validation We validate our demand-system approximation by comparing its predictions with realized price changes in financial markets that arise from actual flow-driven trading. The logic is straightforward: if our approximation captures the right economics, then the patterns it implies under hypothetical reallocations should also emerge in the data when flows occur in reality.

We begin by computing quarterly flows for each active mutual fund, measured as the change in total net assets (TNA) beyond what is explained by the fund's own return,

$$Flow_{j,t} = \frac{TNA_{j,t}}{TNA_{j,t-1}} - (1 + R_{j,t}),$$

where $TNA_{j,t}$ is the total net assets of fund j in quarter t and $R_{j,t}$ is its raw return. Following Lou (2012), these flows are then aggregated into a measure of flow-induced trading for each stock n,

$$FIT_{n,t} = \sum_{j} s_{n,j,t-1}\, Flow_{j,t},$$

where $s_{n,j,t-1}$ denotes the fraction of shares of stock n held by fund j at the end of the previous quarter. Intuitively, $FIT_{n,t}$ captures how much of a stock's trading is mechanically linked to the flows of its fund holders.

To estimate the return effects of these flow-induced trades, we regress quarterly stock returns on $FIT_{n,t}$ while controlling for risk exposures and time effects,

$$ret_{n,t} = \alpha + \gamma\, FIT_{n,t} + \sum_{f} \delta_f\, \beta_{n,t,f} + \lambda_t + \epsilon_{n,t},$$

where $\beta_{n,t,f}$ represents exposures to standard factors such as market, size, and value. The factor loadings are estimated from rolling 60-month regressions (requiring at least 24 months of data), and quarter fixed effects λ_t absorb market-wide shocks. This specification ensures that the estimated coefficient γ isolates the price impact of flows rather than picking up spurious correlations with systematic risk.

Table D.1, Panel A reports the regression results. We estimate a coefficient on FIT of

2.746, indicating that stocks more exposed to flow-induced trading earn higher subsequent returns. The effect is remarkably robust: in a univariate regression without factor controls, the coefficient remains virtually unchanged at 2.737, confirming that the relationship between flows and returns is not simply an artifact of factor exposures.

We then use the estimated $\hat{\gamma}$ to compute the price impact attributable to flow-induced trading,

$$\widehat{ret}_{n,t} = \hat{\gamma}\, FIT_{n,t}.$$

Next, we approximate the corresponding price change using our demand system,

$$\hat{\boldsymbol{\Delta p}}_t = \mathbf{M}_t\, \mathbf{S}_{t-1}\, \mathbf{FIT}_t. \tag{D.1}$$

in which each element n is the predicted price change for stock n, and index m captures cross-elasticity effects.

$$\hat{\Delta} p_{n,t} = \sum_{m=1}^{N} M_{n,m,t} \left(\sum_{j=1}^{J} s_{m,j,t-1}\, \mathrm{Flow}_{j,t} \right), \tag{D.2}$$

We then compare the model-implied price change $\hat{\Delta} p_{n,t}$ with the reduced-form prediction $\widehat{ret}_{n,t}$ in the regression

$$\widehat{ret}_{n,t} = \alpha^* + \xi\, \hat{\Delta} p_{n,t} + \nu_n.$$

Table D.1, Panel B shows that the slope coefficient ξ equals 0.812 with an R^2 of 0.95. In other words, the demand-system approximation and the reduced-form evidence deliver virtually identical conclusions.

This near one-to-one alignment is powerful: it shows that our structural approximation not only replicates the reduced-form correlations emphasized in the literature, but also elevates them by embedding flow effects within a fully specified equilibrium. This structural lens makes it possible to study far more general reallocations—across multiple investor types, assets, and even targeted sectors or funds—while retaining both economic interpretability and predictive power.

Validation from Morningstar Rating Method Change in 2002 In June 2002, Morningstar reformed its rating methodology. Instead of ranking all funds together, it began assigning ratings within "style" categories. As documented by Ben-David et al. (2022), this reform generated exogenous variation in ratings: funds that were upgraded attracted large inflows, while those downgraded experienced outflows. We exploit this natural experiment to provide an additional validation of our approximation.

We first construct a measure of rating changes due to the reform, defined as the difference between a fund's June 2002 rating and its May 2002 rating.[13] Consistent with Ben-David et al. (2022), we find that small value funds experience the largest downgrades, while large growth funds receive the largest upgrades.

To quantify the flow response, we regress Q3 2002 fund flows on these exogenous rating changes:

$$Flow_j = \alpha + \beta\, \Delta ratings_j + controls_j + \epsilon_j.$$

We then isolate the rating-driven component of flows as

$$\widehat{Flow}_j = \hat{\beta}\, \Delta ratings_j.$$

The estimate $\hat{\beta} = 0.0078$ implies that a one-star upgrade increased inflows by roughly 0.78% of fund assets.

Next, using end-of-Q2 holdings, we let $s_{n,j}$ denote the fraction of stock n held by fund j. Treating $\widehat{Flow}_j$ as $\Delta a_{j,t}$ in our equilibrium model, we approximate stock-level price changes in Q3 as,

$$\hat{\boldsymbol{\Delta p}} = \mathbf{M}\,\mathbf{S}\,\widehat{\mathbf{Flow}}.$$

in which element n is the predicted price change for stock n, and index m captures cross-elasticity effects.

$$\hat{\Delta p}_n = \sum_{m=1}^{N} M_{n,m} \left(\sum_{j=1}^{J} s_{m,j}\, \widehat{\mathrm{Flow}}_j \right). \tag{D.3}$$

Stocks more heavily owned by upgraded funds should experience positive price effects, while those held by downgraded funds should decline.

For comparison, we also construct an empirical reduced-form measure. We aggregate the

[13]Because Morningstar ratings are persistent, based on past 2-, 5-, and 10-year performance, the abrupt change from May to June almost entirely reflects the methodology reform rather than performance.

rating-induced flows into stock-level exposure,

$$FIT_n = \sum_j s_{n,j} \, \widehat{Flow}_j,$$

and estimate the return impact directly as

$$ret_n^{(\text{rating-driven})} = \hat{\gamma} \, FIT_n.$$

Finally, we compare our model-implied price changes $\hat{\Delta p}_n$ with these reduced-form return effects by estimating

$$ret_n^{(\text{rating-driven})} = \alpha^* + \xi \, \hat{\Delta p}_n + \nu_n.$$

This exercise provides a demanding out-of-sample test: if our demand-system approximation truly captures the mechanism of flow-driven repricing, it should line up with the quasi-experimental variation induced by Morningstar's 2002 reform.

Table D.1, Panel B reports the results of this comparison. Consistent with the full-sample validation, we obtain a slope coefficient ξ of 0.806 and an R^2 of 0.99, confirming that the demand system's estimated price changes track empirical return movements with striking accuracy.

Table D.1: Validation of Price Approximation Using Flow Shocks

This table validates the demand system's approximation of stock-level price impacts from flow-induced trading. Panel A reports regressions of realized quarterly stock returns on flow-induced trading (FIT), controlling for factor exposures and quarter fixed effects. Panel B compares model-implied price changes from the demand system with empirically estimated return effects, both in the full sample (1981–2019) and in the quasi-experimental 2002 Q3 Morningstar rating reform subsample. Coefficients on ξ close to one and high R^2 values confirm that the demand system closely tracks observed price effects. Parentheses report t-statistics. Standard errors are clustered by stock and quarter in the full sample and are heteroskedasticity-consistent (White, 1980) in the 2002 Q3 subsample.

Panel A: Returns on FIT		
	(1)	(2)
FIT	2.737	2.746
	[8.14]	[7.95]
CAPM Beta		-0.012
		[-3.16]
SMB Beta		-0.011
		[-4.72]
HML Beta		0.007
		[2.16]
Panel B: FIT vs. Demand System Price Effects		
	Full Sample	2002 Q3
ξ	0.812	0.806
	[66.06]	[164.13]
R^2	0.95	0.99