Online Appendix

A Data Description

A.1 Stock Characteristics for Construction of Anomaly Portfolios

We start with the list of 62 stock characteristics from Freyberger et al. (2020). Detailed descriptions and precise variable definitions are provided in the Internet Appendix of that paper, available at https://academic.oup.com/rfs/article-abstract/33/5/2326/5821383. Unless otherwise noted, we adhere strictly to their definitions. We make several modifications to the original list of characteristics, as outlined below.

All the portfolios in Freyberger et al. (2020) are constructed using breakpoints based on stock characteristics at either annual or monthly frequencies. Since we use the 13F quarterly holdings data, we require anomaly portfolio returns at the quarterly frequency. Therefore, we remove several anomaly portfolios which are rebalanced at the monthly frequency. This approach allows us to fit our quarterly holding data and enables us to calculate the counterfactual returns at the desired frequency.

Specifically, we remove:

- Trailing 12-months dividends to price (LDP) from Value category
- Industry-adjusted sales-to assets (SAT_adj) from Profitability category
- All the characteristics from Trading Frictions category except total assets (AT), price times shares outstanding (LME) and industry-adjusted LME (LME_adj). We use the annual breakpoints for construction of LME and LME_adj. We classify these three portfolios under our "Size" category.

We also remove the percentage change in shares outstanding (% Δ Shrout) from the Investment category because the log change in shares outstanding (Δ So) is already included in the Value category. These stock characteristics are highly correlated, and sorting by them produces similar anomaly portfolios.

Second, we do not use the sorts on past returns from Freyberger et al. (2020) due to their requirement for monthly rebalancing. Instead, we consider three alternative momentum portfolios based on the relative strength strategy from Jegadeesh and Titman (1993). The first portfolio is formed using 12-month lagged returns and held for 12 months. Similarly, the second and third portfolios are formed using 6-month and 12-month lagged returns, held for 6 and 3 months, respectively.

Finally, we include several additional anomaly portfolios. The "Issuance" category includes two portfolios sorted on shares issuance from Daniel et al. (2020) using 1-year net-share-issuance (NSI) and 5-year composite-share-issuance (CSI). We also add the Quality-minus-Junk portfolio from Asness et al. (2019) to the "Profitability" Category. Table A.1 summarizes our combined list of 47 stock characteristics used in construction of anomaly portfolios.

A.2 Stock Characteristics for Estimation of Demand System

We include 7 characteristics in Equation (2): log market equity, log book equity, dividends-to-book equity, profitability, investment, market beta and accruals. Below we describe how we define these characteristics.

The first 5 characteristics are defined exactly as in the original Koijen and Yogo (2019)'s approach which closely follows Fama and French (2015). Log market equity is the natural logarithm of price times shares outstanding. Log book equity is the natural logarithm of book equity which is constructed as in Fama and French (2015). Profitability is the ratio of operating profits to book equity. Investment is the annual growth rate of assets.

We calculate the stock's market beta as in Freyberger et al. (2020). Specifically, we follow Frazzini and Pedersen (2014) and define the market beta as the product of correlations between the stock's excess return and the ratio for volatilities. We calculate volatilities using the standard deviations of daily excess returns over a one-year horizon with at least 120 non-missing observations. We estimate the correlations using the overlapping 3-day log excess returns over a 5-year periods with at least 750 non-missing observations.

Finally, we add accruals since our model includes a short-selling sector. We calculate the stock's accrual ratio as in Sloan (1996) where the ratio is defined as the change in non-cash current assets less the change in current liabilities (exclusive of short-term debt and taxes payable) less the depreciation expense, divided by the total assets.

A.3 Holdings Data Construction

We first obtain the Thomson Reuters Institutional Ownership S34 database, covering the holdings of all 13F institutional investors from 1980 to 2019. We use the updated and partially regenerated version of the S34 dataset, which includes corrections to errors previously identified by researchers (e.g. Ben-David et al., 2021).[23] Following Koijen and Yogo (2019), we apply the same filters to the data and the same classification of all 13F investors, categorizing them into six groups: investment advisors, mutual funds, banks, insurance companies, pension funds, and other 13F institutions (e.g., endowments, foundations, and non-financial corporations).

Definition of Short Sellers Unlike Koijen and Yogo (2019), we include short sellers as a seventh investor category, following Mainardi (2021). We obtain firm-level short interest from Compustat's Supplemental Short Interest File and calculate the dollar value of short interest for each stock at the end of each quarter. The assets under management for the short-selling sector are the sum of these dollar holdings. Mainardi (2021) takes a different approach by adding short interest to residual holdings each quarter, which inflates the total number of shares outstanding. This approach creates an additional price gap between actual and counterfactual prices when computing counterfactual prices. To avoid this, we do not follow Mainardi (2021)'s method. Instead, we subtract short-seller ownership from residual ownership while leaving total shares outstanding unchanged. This adjustment likely leads to an underestimation of the effects of residual ownership.

Definition of Small Investors To further classify the remaining ownership, we use the code from Schwartz-Ziv and Volkova (2024) to construct positions held by non-13F blockholders. We download all 13D and 13G filings and their amendments from EDGAR, using fuzzy matching to remove institutional investors whose names are documented in the 13F dataset. These filings occur whenever there are changes in holdings for qualified investors.

[23] See https://wrds-www.wharton.upenn.edu/documents/952/S12_and_S34_Regenerated_Data_2010-2016.pdf for details.

We assume their positions do not change between filing dates and interpolate the quarterly holdings for investors whose 13F records are missing for a few quarters. Since Forms 13D and 13G became mandatory in 1999, our final sample period covers 1999-2022.

We label the remaining ownership—roughly 25–30% of market equity in an average quarter—as "small" investors. By definition, these small investors are not large enough to file 13F, do not hold more than 5% of the stock, and hold long-only positions. This definition differs from most studies using institutional holdings data, which typically attribute all non-13F holdings to a single type of residual investor, usually called "households." Our ownership by small investors is a fraction of household ownership that cannot be attributed to insiders, blockholders, and small non-13F institutions holding large blocks of shares. This distinction makes our measure more precisely reflect the holdings of small individual investors.

Under certain SEC conditions, foreign institutional investors may not be required to file 13F forms, even if they manage more than $100 million in assets. As a result, some of these institutions may be classified under residual ownership. To address this, we examined how sovereign wealth funds (SWFs), one of the most important groups of foreign institutions, are categorized in our dataset. We found that, of the top five SWFs by assets under management (AUM) — Norges, CIC, SAFE, Abu Dhabi, and PIF — only Abu Dhabi does not file a 13F. Among the top 20 SWFs, nine file 13F forms, and six others file alternative forms, allowing us to categorize 15 out of the top 20 SWFs as separate institutions. While it is not possible to capture all foreign institutions using publicly available data, our data collection procedure effectively categorizes the largest SWFs, ensuring minimal residual ownership misclassification.

Table A.1: Description of Characteristics for Construction of Anomaly Portfolios

This table reports the description of the stock characteristics used to construct 47 anomaly portfolios and classify these portfolios into seven broad categories.

Portfolio Label in Freyberger et al. (2020)	Short Description
	A. Value
A2ME	Total assets over market equity
bm	Book equity over market equity
bm_adj	Industry-adjusted book-to-market ratio
C	Cash and short-term investments over total assets
C2D	Cash flow to total liabilities
Δ So	Log change in the shares outstanding
Debt2P	Debt over market equity
e2p	Income over market equity
FCF	Cash flow to book equity
NOP	Net payout ratio
O2P	Payout ratio
Q	Tobin's Q
s2p	Sales over market equity
Sales_g	Sales growth
	B. Momentum
	12-month lagged returns, held for 12 months
	12-month lagged returns, held for 3 months
	6-month lagged returns, held for 6 months
	C. Profitability
ATO	Sales over lagged net operating assets
CTO	Sales over lagged total assets
$\Delta(\Delta$Gm-ΔSales)	Differences in growth of gross margin and sales
EPS	Earnings per share
IPM	Pre-tax income over sales
PCM	Price-to-cost margin
PM	Profit margin
PM_adj	Industry-adjusted profit margin
PROF	Gross profitability over book equity
RNA	Return on net operating assets
ROA	Return on assets
ROC	Return on capital
ROE	Return on equity
ROIC	Return on invested capital
S2C	Sales over cash and short-term investments
SAT	Sales over total assets
	Quality score from Asness et al. (2019)

Table A.2: Description of Characteristics for Construction of Anomaly Portfolios

This table reports the description of the stock characteristics used to construct 47 anomaly portfolios and classify these portfolios into seven broad categories.

Portfolio Label in Freyberger et al. (2020)	Short Description
	D. Investment
Investment	% Change in total assets
Δ ceq	% Change in book equity
Δ PI2A	Change in PP&E over lagged total assets
IVC	Change in inventories over average total assets
NOA	Net operating assets over lagged total assets
	E. Issuance
	1-year net-share issuance
	5-year composite-share issuance
	F. Size
AT	Total assets
LME	Market equity
LME_adj	Industry-adjusted market equity
	G. Tangibility
AOA	Absolute value of operating accruals
OL	Operating leverage
OA	Operating accruals

B Nested Logit and Related Tests

B.1 Industry-level Nested Logit

B.1.1 Why We Nest by Industry

In theory, the cleanest way to relax the simple-logit assumption would be to nest by factors, allowing investors to switch across anomaly portfolios. In practice, this approach is infeasible. The 47 long–short anomalies overlap heavily, so "moving from one factor to another" largely reflects a reshuffling of shared positions rather than a move to a distinct investment style. Because the portfolios are not mutually exclusive, any factor-level nest would violate the independence assumptions of the demand system and generate ill-defined cross-elasticities.

Industry nesting provides a practical solution. Each stock belongs to exactly one industry, which eliminates attribution problems—for example, determining the extent to which a stock simultaneously contributes to both value and momentum—while preserving a meaningful economic margin along which investors reallocate capital. It is the tightest substitute group we can model without violating the structure of the data.

The concentration metrics underscore that anomaly portfolios are far from industry-neutral. With the Fama–French ten-industry classification, an equal-weighted benchmark (10 industries $\times$ 10% each) implies an HHI of 10%. By contrast, the mean Herfindahl–Hirschman Index across our 47 anomalies is 22% (Table B.1)—more than double the neutral benchmark—demonstrating that these strategies are materially tilted toward a subset of industries rather than evenly diversified across all ten sectors.

When we collapse the ten industries into five broader "super-industries," the mean HHI climbs to roughly 31% (Table B.2) against the industry-neutral benchmark of 20%. Such concentration makes within-industry rotation the immediate and quantitatively important substitution channel, giving an empirical justification for our two-tier (industry-first, stock-second) nested-logit specification.

B.1.2 Why We Stop at Five Industries

Choosing the outer nest is a balancing act: more industries sharpen substitution estimates but starve the data that pin down each investor's demand curve. Recall that we follow KY and retain any institution with more than 1,000 holdings as a stand-alone investor. All smaller accounts are pooled by investor type and AUM until each pool spans at least 2,000 stocks. However, because we estimate demand curves for both the pooled blocks and the individual investors (per our Ridge procedure), we need to ensure that every investor—pooled or not—retains a sufficient number of observations for reliable estimation.

When we move the estimation to the industry level, the number of stocks per investor drops fast. Even if holdings were uniformly distributed, dividing the universe into five industries would leave each large investor with roughly 200 stocks per industry cell, and only about 100 when using ten industries. In reality, portfolios are industry-concentrated, so several sectors fall well below these averages. Once the per-industry sample size falls into the low double digits, the nested-logit likelihood becomes ill-conditioned, and the estimation fails to converge.

Table B.3 illustrates what happens when we choose the Fama–French 10-industry grid. Even among the largest institutions (Panel A), the average number of stocks in thin sectors such as Durables or Telecom often falls below 50. For example, large banks hold an average of only 47 Durables stocks and 46 Telecom stocks. The minima are even lower: the same banks sometimes hold as few as 23 Durables stocks, and investment adviser stock count drops to just 19 stocks in that cell. Panel B reveals that this issue is even more pronounced for smaller investors, with minimum stock counts falling into the low double digits across several industries. With counts this low, price–elasticity estimates for those industries would hinge on a handful of securities.

Collapsing the ten sectors into the five "super-industries" reported in Table B.4 substantially increases cell sizes. For large banks, the minimum stock count now exceeds 87 in every industry, and for small banks the mean never falls below 77. Among small insurance companies and pension funds, the mean stock count in their sparsest industry remains above 80. The thicker cells supply the variation needed to identify industry-level elasticities without facing convergence problems.

Finally, the loss of economic detail is modest. The industry-concentration numbers reported earlier indicate that the five-way split still captures most of the concentration evident in the finer grid. Taken together, these results indicate that the FF-5 is the narrowest classification that preserves statistical power while retaining the substitution margin most relevant for investors. We therefore adopt FF–5 as the baseline throughout the nested-logit analysis.

B.1.3 Outside Assets and Investor Grouping

The outer nest requires an outside option that is present—and of comparable scale—in every industry. Using KY's definition alone (REITs, foreign listings, missing-characteristic stocks) fails that test: Table B.5, Panel A shows outside weights that cluster in the residual "Other" sector while falling below one percent in several core industries.

Such imbalances would let a handful of industries absorb nearly all substitution into the outside asset and may destabilize the price-elasticity estimates. The outer nest treats each industry as a competitor for the investor's dollar, so every industry must contain a credible "last resort" security that absorbs capital when the tracked names in that sector lose their appeal. If the outside sleeve is large in one industry but nearly absent in another (as shown in Panel A), an investor wishing to step away from, for example, Manufacturing faces two very different alternatives: she can exit into a sizeable Manufacturing outside asset or reallocate to Health Care, whose small outside sleeve offers no comparable refuge. That asymmetry biases the estimated nesting parameter toward industries that happen to carry a disproportionately large outside weight.

We restore balance by adding the Koijen et al. (2023) (KRY) tail filter. Within each industry, we reassign the smallest x% of market capitalization to the outside bucket and then append the standard KY outside assets. At x=10% (illustrative) the result is a much smoother distribution: Panel B shows outside weights tightly clustered between 0.1% and 1% across all industry–investor pairs. Each sector now offers a comparably sized "escape hatch," so withdrawing from Health Care into its outside assets is mechanically no easier—or harder—than withdrawing from Manufacturing or HiTec. This symmetry stabilizes the outer-nest likelihood and yields price-elasticity estimates that are not driven by a single, disproportionately large outside pool.

This balanced construction also sharpens economic interpretation. Because every industry supplies its own minimal outside option, our nested logit can cleanly distinguish two actions: reallocating *within* an industry (into its outside sleeve) versus reallocating *across* industries (changing the outer weight). Without the KY+KRY blend, those two moves would blur together whenever an investor fled a sector whose outside sleeve was effectively empty. We therefore adopt the combined definition in all subsequent work and fine-tune the threshold x below to maximize sample size while keeping the outside shares rich enough and uniformly distributed.

Balancing the outside bucket across sectors is only half the problem; institutions must also *hold* enough of those securities for the inner-nest logit to identify a price effect. Table B.6 makes the shortfall clear. Although mean counts appear adequate—the average bank, for example, holds nine to seventeen outside stocks in most sectors—the lower tail collapses. A 20th-percentile bank holds fewer than one HiTec outside stock, and the 20th-percentile investment adviser holds none at all in two of the five industries. A logit with zero or one observation in the outside category reduces to a corner solution and injects noise into the elasticities.

To address this issue, we drop investors with effectively no outside positions and transfer their assets to small investors. Table B.7 shows that this pruning minimally distorts the institutional landscape. Even under the strictest cutoff, x=1%, the reclassified AUM is trivial for banks, insurers, mutual funds, and pensions (all below 7 percent), and peaks at only about 22 percent for investment advisers. Increasing x further has little effect on these fractions but sharply reduces the number of stocks available for the inner logit, thereby weakening statistical power.

We therefore fix x=1%. This choice maximizes the stock counts available for price identification while keeping institutional reallocations negligible. After these adjustments every remaining investor–industry cell contains a meaningful outside sleeve, allowing the nested logit to cleanly disentangle within-industry exits from cross-industry reallocations.

B.1.4 Comparison Between Industry-level Nested Logit and Koijen and Yogo (2019)'s simple logit

Investor–Industry Substitution Elasticities Table B.8 shows that industry membership is the dominant substitution margin. Across investor categories, the Technology and Health-Care nests record the highest elasticities, with averages of 0.52 and 0.45—roughly double the value observed for the "Other" nest (0.21). These elevated elasticities imply that capital exits growth-oriented sectors swiftly whenever their inclusive values deteriorate. Energy, Utilities, and other heavily regulated sectors sit at the opposite end of the spectrum (average $\sigma \approx 0.18$), consistent with trading frictions and information gaps. The five-fold inter-industry dispersion in σ underscores the need to let each nest carry its own inclusive value; collapsing all industries into a single outer nest would mask the substitution behavior.

Investor heterogeneity remains economically meaningful. Averaged across industries, small investors exhibit the highest elasticity ($\bar{\sigma} = 0.78$), reflecting their limited trading constraints. Mutual funds and investment advisers follow at $\bar{\sigma} \simeq 0.35$–0.50, consistent with benchmark-driven rebalancing. Banks and insurance companies exhibit elasticities of approximately $\bar{\sigma} \simeq 0.25$ and $\bar{\sigma} \simeq 0.39$, respectively, as risk-based capital rules and asset–liability matching dampen their willingness to substitute across assets. Short sellers are almost inelastic outside Consumer stocks, mirroring scarcer borrow supply in Technology and Health. These results confirm that the nested specification preserves investor-specific frictions.

The interaction between industry and investor characteristics is even more revealing. Banks, for example, raise their elasticity from 0.15 in "Other" industries to 0.40 in Health Care—a 160% increase—while the most flexible retail investors see their elasticity drop by nearly 50% when moving from Technology (0.88) to regulated sectors (0.47). Industry attributes therefore amplify or dampen every investor's propensity to reallocate. This complementarity underscores the importance of industry nests and drives the variance-decomposition and counterfactual analyses that follow.

Stock–Level Elasticities Table B.9 compares stock–level price elasticities obtained from the baseline KY-based methodology (Panel A) with those from the nested-logit specification

(Panel B). Averaging across all stocks within each investor group, the nested model produces higher elasticities for every institutional category: bank elasticities rise from 0.08 to 0.14, insurer elasticities from 0.13 to 0.21, adviser elasticities from 0.42 to 0.54, and pension-fund elasticities from 0.10 to 0.17. Panel C confirms that these increases are significant, with t-statistics between 5 and 15. Small investors are the only exception—their mean elasticity edges down by 0.02 and the difference is statistically indistinguishable from zero—yet their elasticity remains an order of magnitude above that of more constrained institutions.

Why does nesting raise elasticities? The simple logit treats every stock as an equally plausible substitute, so price pressure on one name is partially absorbed by migration to the entire market. The nested structure first allocates capital across industry nests, then across stocks within a chosen nest. When a stock moves in price, substitution is now largely confined to its own industry, a narrower opportunity set that flattens the demand curve and yields a larger price elasticity. This mechanical effect is strongest for investors whose holdings cluster within a few sectors—banks, insurers, and pension funds—precisely where the elasticity jump is largest. Small investors as a group already hold a broad portfolio of stocks, so limiting substitution to the industry level does little to alter their marginal response.

These patterns speak directly to the concern that the homogeneous cross-elasticity inherent in the simple logit biases institutional elasticities downward when clear substitution channels exist (Chaudhary et al. (2025); Fuchs et al. (2025)). Allowing industry nests relaxes the restrictive logit structure and realigns our elasticity estimates with the theoretical intuition that sector-bound mandates curtail investors' ability to reallocate across the full universe of stocks. The 60–120 bp increases for banks, insurers, and advisers attest that the simple logit model indeed understated their sensitivity to price movements.

Portfolio–Level Elasticities Table B.10 shows that moving from individual stocks to anomaly portfolios dampens price sensitivity, but the decline is smaller once industry nests are introduced. In the simple logit model, the dollar-weighted stock elasticity averages about 0.46, while aggregation to long–short portfolios reduces it to roughly 0.30. By contrast, the nested-logit elasticities fall only to the 0.38–0.43 range. When substitution is restricted to economically relevant groups—in our case, industries—price changes cannot be fully offset by reallocating to unrelated names. As a result, the demand curve remains flatter, and elasticities stay higher at both the stock and portfolio levels of aggregation.

Panel A illustrates the point anomaly by anomaly. Take the Value strategy: under the simple logit model the long (short) leg has an elasticity of 0.31 (0.29); nesting lifts these figures to 0.41 and 0.38. Similar ten-basis-point increases appear in every other anomaly, and the symmetry between long and short legs is preserved, indicating that nesting does not mechanically bias one side of the trade.

Panel B extends the comparison to all quintiles across the 47 characteristic sorts. The nested specification adds 0.09–0.10 to the elasticity in every portfolio slice with t-statistics above 10. The uplift is therefore pervasive, not driven by a handful of headline portfolios.

These findings address the concern that the homogeneous cross-elasticity of the simple logit artificially suppresses price sensitivity when clear substitution patterns exist. Industry nesting corrects that bias: it delivers higher, more realistic elasticities while leaving the cross-anomaly ranking intact.

B.1.5 Newton Algorithm for Counterfactual Prices

This section details the Newton scheme we develop to solve the fixed-point $\boldsymbol{p} = f(\boldsymbol{p})$ for the calculation of counterfactual prices in the nested logit model. The algorithm generalizes the Koijen and Yogo (2019) procedure to the nested-logit setting and preserves its key advantages: fast convergence, minimal memory footprint, and robustness to ill-conditioned Jacobians.

Mapping Definition and Economic Intuition We aggregates investors' optimal holdings into market demand and use the standard "supply-equals-demand" equilibrium condition which clears at the equilibrium price,

$$\boldsymbol{D}(\boldsymbol{p}) = \sum_{i=1}^{I} A_i\, \bar{w}_i(\boldsymbol{p}), \qquad \boldsymbol{p} = f(\boldsymbol{p}) = \log \boldsymbol{D}(\boldsymbol{p}) - \boldsymbol{s},$$

where $\boldsymbol{s}$ is the log supply vector. The mapping $f(\boldsymbol{p})$ pushes current prices toward the log market-clearing prices implied by investors' nested demands. A fixed point therefore represents a competitive equilibrium under the estimated preference parameters.

Newton Update Starting from an initial guess $\boldsymbol{p}_0$, we iterate

$$\boldsymbol{p}_{m+1} = \boldsymbol{p}_m + \boldsymbol{J}_m^{-1}\big(f(\boldsymbol{p}_m) - \boldsymbol{p}_m\big), \qquad \boldsymbol{J}_m = \boldsymbol{I} - \frac{\partial f(\boldsymbol{p}_m)}{\partial \boldsymbol{p}^\top}.$$

Because f is a contraction near the solution, $\boldsymbol{J}_m$ is strictly diagonally dominant; its inverse therefore exists and can be computed efficiently.

Jacobian Approximation The exact $\frac{\partial f(\boldsymbol{p}_m)}{\partial \boldsymbol{p}^\top}$ is dense and expensive to invert. Following KY we keep only its diagonal:

$$\frac{\partial f(\boldsymbol{p}_m)}{\partial \boldsymbol{p}^\top} \approx \operatorname{diag}\Big(\min\Big[\tfrac{\partial f(\boldsymbol{p}_m)}{\partial p_n}, 0\Big]\Big)_{n=1}^{N},$$

which (i) preserves the Newton direction, (ii) guarantees that $\boldsymbol{J}_m^{-1}$ is trivial to compute, and (iii) prevents overshooting by clipping positive diagonal entries to 0.

Gradient Computation We rewrite the diagonal entries as

$$\frac{\partial f(\boldsymbol{p}_m)}{\partial p_n} = \frac{1}{\sum_i A_i\, w_i(\boldsymbol{p}_m)} \sum_{i=1}^{I} A_i \frac{\partial w_i(n)}{\partial p_n}.$$

Applying the chain rule to $w_i(n) = \bar{w}_i(\gamma)\, w_i(n|\gamma)$ yields two terms:

$$\frac{\partial w_i(n)}{\partial p_n} = A_{i,n} + B_{i,n},$$

with

$$A_{i,n} = \bar{w}_i(\gamma)\big(1 - \bar{w}_i(\gamma)\big) w_i^2(n|\gamma)\, \sigma_{i,\gamma}\, \beta_{0,i,\gamma},$$

$$B_{i,n} = \bar{w}_i(\gamma) w_i(n|\gamma)\big(1 - w_i(n|\gamma)\big)\, \beta_{0,i,\gamma}.$$

Term $A_{i,n}$ captures the outer-nest spillover generated by industry-level substitution; term $B_{i,n}$ reproduces the usual inner-nest sensitivity. Both are cheap to evaluate because they reuse quantities already computed for the demand shares.

B.2 Reduced-form Analyses of Substitution in Equity Portfolios

To verify that the industry nest we impose is not merely a statistical convenience but the margin along which investors actually reallocate capital, we turn to the reduced-form "flow–multiplier" test of Chaudhary et al. (2025) (CFL). The approach regresses returns on demand shocks proxied with surprises in flow-induced trading and then sequentially adds fixed effects to observe how quickly the multiplier collapses as potential substitutes are absorbed. If within-industry rotation is indeed the dominant channel—as assumed by the nested logit—introducing industry dummies should materially reduce the flow coefficient. This exercise provides a model-free validation of whether industry-level substitution is the key margin along which investors reallocate capital in the data.

CFL show that a log-linearized nested-logit model collapses to a simple reduced form in which two parameters fully describe price responses. The own multiplier M measures the percentage increase in an asset's price in response to a 1% exogenous increase in investor demand for that asset. The substitute pass-through coefficient $\widetilde{M}$ captures the percentage increase in the asset's price when the prices of its close substitutes rise by 1%.

Consider a representative investor facing groups g (in our setting, industries):

$$\begin{aligned}
w(j \mid g) &= \frac{\exp\{\delta(j,g)\}}{\sum_{j \in g} \exp\{\delta(j,g)\}}, \\
w(g) &= \frac{\left[\sum_{k \in g} \exp\{\delta(k,g)\}\right]^{\lambda}}{1 + \sum_{g'} \left[\sum_{k \in g'} \exp\{\delta(k,g')\}\right]^{\lambda}}, \\
\delta(j,g) &= \beta_g p_j + \beta X + u_j.
\end{aligned} \tag{B.1}$$

The nesting parameter λ controls substitution at the group level; $\lambda = 1$ reproduces the simple logit.

Log-linearising (B.1) and imposing market-clearing yields

$$\Delta p_j = \underbrace{\frac{1}{1-\beta_g}}_{M} u_j + \underbrace{\frac{\beta_g}{\beta_g - 1}(1-\lambda)}_{\widetilde{M}} \underbrace{\sum_{k \in g} w(k \mid g)\, \Delta p_k}_{\Delta p_j^{\text{sub}}} + \underbrace{\lambda \frac{1}{\beta_g - 1} \sum_{g'} \sum_{k \in g'} \beta_{g'} w(k)\, \Delta p_k}_{\text{time FE}} + \nu_j. \tag{B.2}$$

With an instrument for the demand shock u_j and for the substitute-portfolio return Δp_j^{sub}, equation (B.2) can be estimated in a single regression. In our context each group g is one of the five Fama–French industries, so $\widetilde{M}$ directly tests how tight within-industry substitution really is: higher values signal that shocks propagate quickly to its industry peers, consistent with the nesting structure we impose in the structural model. Under homogeneous substitution across all assets (i.e., $\lambda = 1$), the structural model implies $\widetilde{M} = 0$. In contrast, larger values of $\widetilde{M}$ indicate stronger within-group comovement, consistent with heterogeneous substitution patterns as captured by the nested logit structure introduced in the prior sections.

B.2.1 Construction of Flow Shocks

Our instrument follows the four–step protocol in CFL, Sect. 3.3, omitting only the latent-factor purge to align with their baseline specification. Throughout, i indexes U.S. domestic equity mutual funds drawn from CRSP/Thomson, j denotes CRSP common stocks, and t is a fiscal quarter.

(i) Raw percentage flow. For each fund we remove capital gains from end-of-quarter AUM to obtain net investor contributions,

$$F_{i,t} \; = \; \mathrm{AUM}_{i,t}^{\mathrm{end}} - \mathrm{AUM}_{i,t-1} - r_{i,t}\mathrm{AUM}_{i,t-1},$$

where $r_{i,t}$ is the fund's net return. Dividing by beginning-of-quarter assets gives the raw percentage flow

$$f_{i,t} \; = \; \frac{F_{i,t}}{\mathrm{AUM}_{i,t-1}}. \tag{B.3}$$

We winsorise $f_{i,t}$ at the 1st/99th percentiles to guard against stale or mis-recorded AUM figures.

(ii) Idiosyncratic innovation and long-run impact. Serial correlation is removed fund-by-fund with an AR(3) plus linear trend,

$$f_{i,t} = \alpha_i + \rho_1 f_{i,t-1} + \rho_2 f_{i,t-2} + \rho_3 f_{i,t-3} + \delta_i t + u_{i,t}, \tag{B.4}$$

leaving $u_{i,t}$ as the one-quarter innovation. The cumulative effect of that shock under the estimated persistence pattern is

$$\kappa_i = \frac{1}{1 - \rho_1 - \rho_2 - \rho_3}. \tag{B.5}$$

(iii) Double-demeaning. To remove fund fixed effects driven by marketing channels (μ_i) and common liquidity waves (λ_t) we write

$$u_{i,t} = \mu_i + \lambda_t + \tilde{u}_{i,t},$$

and keep the residual $\tilde{u}_{i,t}$ as our idiosyncratic, time-varying flow shock.

(iv) Aggregating to the stock level. Let $\theta_{ij,t-1}$ be the fund i's fraction of stock j' shares held by all the U.S. domestic equity mutual funds at $t-1$. Under proportional trading the percentage change in fund i's holding of stock j induced by the flow innovation is $\theta_{ij,t-1}\tilde{u}_{i,t}$.[24] Summing over all funds and scaling by κ_i produces the stock-level shock

$$\mathcal{F}_{j,t} = \sum_i \theta_{ij,t-1}\,\kappa_i\,\tilde{u}_{i,t}, \tag{B.6}$$

interpreted as the predicted *percentage* change in mutual-fund ownership of stock j that arises purely from unexpected flows.[25]

(v) Aggregating to the portfolio level. For any anomaly or industry portfolio P we form a flow shock by value-weighting the constituent stock shocks. Using lagged market-value weights $w_{j,P,t-1}$, the portfolio-level shock is

$$\mathcal{F}_{P,t} = \frac{\displaystyle\sum_{j\in P} w_{j,P,t-1}\,\mathcal{F}_{j,t}}{\displaystyle\sum_{j\in P} w_{j,P,t-1}}, \tag{B.7}$$

i.e., $\mathcal{F}_{P,t}$ represents the predicted percentage change in mutual-fund ownership of the entire portfolio that is driven solely by unexpected flows into its component stocks. We use (B.7) when estimating the CFL regression at the anomaly-portfolio level and when instrumenting the substitute return for each industry.

[24] We scale holdings by total shares held by all the mutual funds rather than by total shares outstanding to be consistent with CFL and Wardlaw (2020).

[25] We include a time trend in (B.4). The (time-series average of the cross-sectional) correlation between stock-level (surprise in) flow-induced trading $\mathcal{F}_{j,t}$ constructed with and without the time trend is ≈ 0.8.

B.2.2 Stock-level CFL Regressions

To examine the industry margin, we start by estimating the reduced-form equation (B.2) for individual stocks, replacing the model-generated substitute portfolio $\Delta p_{j,t}^{\text{sub}}$ with interacted fixed effects of the form Industry $\times$ Quarter. This follows CFL's preferred implementation and let the data decide how much of an own-shock propagates to other stocks once we absorb common industry–time variation.

Column (1) of Table B.11 presents the baseline regression of stock returns on stock-level demand shocks, controlling only for time fixed effects. This specification implicitly assumes homogeneous substitution across all stocks. The coefficient on the flow shock $\mathcal{F}_{j,t}$ is $M \simeq 0.175$ indicating that a 1% demand shock raises the stock price by 17.5 basis points.

In Column (2), we introduce group-time fixed effects using the Fama-French 5-industry classification to control for comovement with close substitutes. The estimated multiplier declines slightly to 0.158, a decline of about ten percent. Expanding the partition to ten and forty-nine Fama–French industries—Columns (3) and (4)—has almost no additional impact; the multiplier stabilizes at 0.153.

The modest attenuation of M confirms the main lesson of the CFL's equity appendix: at the *stock* level, cross-stock substitution play a limited role. Equally important, the small difference between the FF–5 and FF–49 specifications indicates that our coarser FF–5 nesting is already sufficiently granular to capture the relevant within-industry substitution.

Taken together with the structural elasticity estimates, these reduced-form results reinforce our choice of a five-industry outer nest. Industry dummies absorb the key substitution patterns without overturning the multipliers, and finer partitions offer no material improvement.

For comparison, the analogous specification in CFL yields a multiplier of roughly 0.30. Direct comparability is limited because the two studies draw on different mutual-fund holdings databases (Morningstar versus CRSP) and cover distinct sets of stock–quarter observations (about 145,000 in CFL versus 226,000 here), but the methodological setups are otherwise aligned. CFL report an almost identical pattern of limited attenuation — their stock-level multiplier falls only slightly once industry-by-time fixed effects are introduced. This close correspondence reinforces the conclusion that, at the *stock* level, cross-stock substitution within an industry only modestly dampens the price impact of mutual-fund flows, regardless of the underlying holdings database.

B.2.3 Portfolio-level CFL Regressions: Industry Margin

We repeat the CFL reduced-form exercise for the ~ 100 assets that feed our variance decompositions.[26] At this level, the "own" shock is the flow-induced return on a portfolio P, rather than on an individual stock, and the key question is whether industry rebalancing still absorbs most of the flow pressure once portfolios span multiple sectors.

Introducing full Industry×Quarter fixed effects—our strategy at the single-stock level—breaks down in the portfolio setting for two reasons. Conceptually, a fixed effect treats each observation as if it belonged to a *single* sector, absorbing *all* variation that is common to stocks in that sector during the quarter. An anomaly portfolio, however, is deliberately diversified: it often carries 20–30% of its weight in one industry, 20-30% in a second, and non-trivial stakes in a third or fourth. Forcing such a portfolio into a single-sector dummy is therefore arbitrary and discards the exposure information carried by its other industries—information that is essential for measuring substitution. Practically, the data are thin: with only about one hundred buckets per quarter, saturating the regression with five (or ten) industry dum-

[26] Since we consider 47 signals and both long and short sides, we work with a total of 47×2 assets.

mies interacted with 94 time periods would consume more than half the degrees of freedom, inflate standard errors, and render the multiplier imprecise.

To model the sector mix, we replace fixed effects with an *exposure-weighted industry return*. We construct a continuous industry-substitute control that matches each bucket's actual sector mix. Let P denote a long or short leg of an anomaly-based strategy, k denote a given industry out of $Nind$, and define the industry-substitute portfolio as:

$$s^{\text{Ind}}_{P,k,t} = \frac{\sum_{j \in P \cap k} \text{MktCap}_{j,t}}{\sum_{j \in P} \text{MktCap}_{j,t}} \quad \text{and} \quad \mathcal{R}^{\text{Ind}}_{P,t} = \sum_{k=1}^{Nind} s^{\text{Ind}}_{P,k,t} \, R_{k,t}, \tag{B.8}$$

where $R_{k,t}$ is the Fama–French industry return. We consider Fama-French 5-, 10- and 49-industry classifications (so $Nind = 5, 10$ or 49). This single regressor preserves the portfolio's true, value-weighted industry mix and scales in direct proportion to that exposure: if a bucket doubles its HiTec weight, the sensitivity of $\mathcal{R}^{\text{Ind}}_{P,t}$ doubles automatically. It thus retains continuous cross-sectional variation while consuming only one degree of freedom. Economically, $\mathcal{R}^{\text{Ind}}_{P,t}$ captures the exact margin of within-industry substitution posited by our nested logit, while leaving room for the regression to detect additional cross-industry reallocation. The result is a tighter, more interpretable test of whether industry rotation absorbs flow pressure at the portfolio level.

Table B.12 begins with the benchmark specification that contains only quarter fixed effects. Column (1) delivers an own multiplier of 2.54. Such a large coefficient translates into a steep demand curve—i.e. very low price elasticity at the portfolio level—and is an order of magnitude above the 0.17 we measured for individual stocks. This gap mirrors Li and Lin (2025)'s insight that aggregation amplifies demand pressure: when the shock hits a portfolio rather than a single ticker, it elicits a commensurately bigger price response.

Because no other control is in the regression, that 2.54 multiplier reflects *both* the direct pressure on the bucket's own holdings and any price movement that spills over from stocks with a similar industry profile.

Column (2) adds our exposure-weighted return based on Fama-French five industries classification, $\mathcal{R}_{P,t}^{\text{Ind}}$, to the regression:

$$r_{P,t} = c + M\mathcal{F}_{P,t} + \widetilde{M}\mathcal{R}_{P,t}^{Ind} + \upsilon_{P,t} \tag{B.9}$$

The own-price multiplier drops sharply from 2.54 to 1.56, a 40% reduction, while the coefficient on the substitute return is 1.80 and highly significant. This variable captures the part of the portfolio's price that comoves with its industry peers. The 40% decline in the own multiplier indicates that roughly two-fifths of the original price response was not specific to the portfolio itself but it was shared with the stocks that sit in the same industries.

The coefficient of 1.80 on $\mathcal{R}_{P,t}^{\text{Ind}}$ implies that a 1% increase in the return of industry peers—e.g., due to mutual fund inflows into similar stocks—raises the portfolio return by approximately 1.8%. The fact that this industry coefficient is larger than the *residual* own multiplier (1.56) suggests that most of the flow pressure is absorbed through trades in the same industries, rather than through unrelated sectors.

Columns (3) and (4) refine the industry control to the FF10 and FF49 grids. The own multiplier edges down to 1.37 and 1.17, respectively—further evidence of intra-industry substitution, but the incremental gain is modest. Similarly, the passthrough coefficient declines in parallel, and the regression R^2 creeps up from 0.930 to 0.954, confirming that ever-finer sector cuts add only marginal explanatory power.

Taken together, the results line up with the nested-logit structure. Most of the elasticity appears once we control for the five super-industries; dividing those industries more finely polishes the fit but does not overturn the message. Investors find it relatively easy to substitute into stocks *within* the same industry, which is why controlling for industry returns removes a large slice of the flow effect; substituting across industries is harder, so a substantial multiplier remains even after very fine sector controls are included. This hierarchy of elasticities is what our five-industry outer nest is designed to capture.[27]

B.2.4 Portfolio-level CFL Regressions: Category and Industry Margins Together

Extension to Category-Level Substitution. Following the spirit of CFL's Table 2, we extend our substitution tests beyond the industry dimension and recognise that investors can also rotate capital across anomaly categories (or broader styles) —Value, Momentum, Profitability, Investment, Tangibility, Issuance, and Size. Treating these seven categories as mutually exclusive substitutes is problematic, however, because anomaly portfolios are built from highly overlapping stock sets. Growth stocks that appear in the short leg of Value anomalies, for example, often re-appear in the long leg of Momentum anomalies, while high-profitability firms heavily populate both the Profitability and Investment portfolios. This overlap creates two econometric challenges. First, any category-level shock (say, a sudden preference for growth firms) is mechanically transmitted to multiple anomaly returns. Second, the degree of overlap varies widely across portfolio pairs, so a blanket fixed-effect that treats all anomalies inside a category as perfect substitutes can over-correct, erasing genuine

[27] We deliberately keep the substitute–return term $\mathcal{R}^{\text{Ind}}_{P,t}$ in levels at this stage, mirroring the "first–pass" descriptive regressions in CFL. Our goal here is simply to diagnose where the excess multiplier originates—whether from own-holding pressure or from comovement with industry peers—before introducing additional econometric structure. In the next section, we instrument all substitute returns and report that the qualitative ranking of coefficients—and hence the economic message—remains unchanged.

cross-sectional information. The remainder of this section sets up two complementary strategies that solve these problems: (i) category × quarter fixed effects that wipe out common category shocks, and (ii) an overlap-weighted substitute-return approach that mirrors our industry-control construction while scaling each substitute by the actual stock overlap with the target portfolio.

We implement two complementary strategies to control for style-level substitution:

1. **Category × Quarter Fixed Effects**: We include category-by-quarter dummies $\delta_{s,t}$ to sweep out broad style movement where $s \in \{$ Value, Momentum, Profitability, Investment, Tangibility, Issuance, Size$\}$. These fixed effects absorb shocks that are common to all portfolios within a style in a given quarter. The associated regression is:

$$r_{P,t} = \alpha + M\mathcal{F}_{P,t} + \widetilde{M}\delta_{s(P),t} + v_{P,t} \tag{B.10}$$

   This approach fully removes any style-wide shocks (e.g., affecting both book-to-market and sales-to-price portfolios in the Value category). However, it imposes the assumption that all anomalies within a style are perfect substitutes, regardless of the actual stock overlap between them. As a result, portfolios with negligible overlap may be treated as perfect substitutes.

2. **Overlap-Weighted Substitute Shock**: This approach mirrors our methodology for estimating the industry-substitution channel and explicitly accounts for stock overlap between portfolios. For any ordered pair of portfolios P, Q, we define the overlap weight:

$$w_{P,Q} \;=\; \frac{\text{mkt-cap of } (P \cap Q)}{\text{mkt-cap of } P}, \qquad 0 \leq w_{P,Q} \leq 1. \tag{B.11}$$

We then construct two versions of the substitute return $\mathcal{R}_{P,t}^{Style}$ based on the set of portfolios Q included in the sum:

(a) **Within-Style Substitute Portfolio**: Limit Q to portfolios within the same style family as P (e.g. only other anomaly portfolios wthin Value), excluding portfolio P itself. This captures substitution within a style category—arguably where switching is most plausible:

$$\mathcal{R}_{P,t}^{sub} = \sum_{Q \in \text{Style(P)}} w_{P,Q}\, R_{Q,t}, \quad \text{(within-style substitute portfolio)} \tag{B.12}$$

(b) **All-anomaly Substitute Portfolio**: Let Q range over all portfolios in the anomaly universe. This broader measure allows for the possibility that investors may switch across styles if there is sufficient stock overlap:

$$\mathcal{R}_{P,t}^{sub} = \sum_{Q \in \text{all anomaly}} w_{P,Q}\, R_{Q,t}, \quad \text{(all-anomaly substitute portfolio)} \tag{B.13}$$

The final regression is:

$$r_{P,t} = \alpha + M\mathcal{F}_{P,t} + \widetilde{M}\mathcal{R}_{P,t}^{sub} + v_{P,t} \tag{B.14}$$

Both specifications allow us to move beyond fixed effects by explicitly accounting for the intensity of stock-level overlap. The within-style version is more targeted and interpretable, while the all-anomaly version serves as a more conservative control.

Table B.13 extends the portfolio–level CFL exercise to account for heterogeneous substitution patterns at the anomaly-style level. We begin by reproducing the benchmark results from Table B.12. Column (1) includes only quarter fixed effects and yields a multiplier of 2.54 on the portfolio-level demand shock. In Column (2), we add the exposure-weighted Fama-French 5-industry return, $\mathcal{R}^{\text{Ind}}_{P,t}$. This inclusion reduces the own-price multiplier to 1.56 and introduces a highly significant industry passthrough coefficient of 1.80. These two columns establish the reference point against which the style–substitution controls are evaluated.

Column (3) introduces Category×Quarter fixed effects, which absorb any shocks common to all portfolios within a given style—such as Value, Momentum, Profitability, Investment, Tangibility, Issuance, or Size. The own multiplier and the industry passthrough shift only slightly, to 1.58 and 1.81, respectively. This suggests that while broad style-level movements account for some additional variation, the dominant flow effect continues to operate through industry-level comovement.

Column (4) adds an overlap-weighted style substitute return. For each bucket we build $\mathcal{R}^{\text{sub}}_{P,t}$ by weighting the returns of other portfolios within the same style by their actual stock-level overlap. The coefficient on this overlap-weighted style control is 0.18 and statistically significant at the 1% level. The own-price multiplier decreases slightly from 1.575 to 1.554. This indicates that allowing the strength of the style control to vary with actual portfolio overlap captures some additional passthrough, consistent with investors reallocating across similar anomaly strategies. However, the modest change in the multiplier suggests that most of the initial flow-induced price pressure had already been accounted for through industry-level substitution.

The final block of the table treats both substitute returns as potentially endogenous and instruments each with its own flow shock. Columns (5) and (6) report the first–stage regressions that gauge the strength of our two instruments. When the dependent variable is the overlap–weighted anomaly substitute return $R_{\text{sub},P,t}$ (column 5), the flow shock to the overlapping portfolios $F^{\text{sub}}_{P,t}$ loads at 5.46 (s.e. 1.45), while the industry flow shock enters at a much smaller 0.46 (s.e. 0.07). Flipping the dependent variable to the exposure-weighted industry return $R^{\text{Ind}}_{P,t}$ (column 6) produces the mirror image: $F^{\text{Ind}}_{P,t}$ carries a coefficient of 1.13 (s.e. 0.10) and $F^{\text{sub}}_{P,t}$ a similar 1.35 (s.e. 0.46). The Cragg–Donald F-statistic of 234 comfortably exceeds conventional weak-instrument thresholds, indicating that the joint set of instruments has ample strength for reliable 2SLS inference.

Column 7 presents the two-stage least-squares estimates that treat *both* substitute returns as endogenous. The own multiplier remains virtually unchanged at 1.53 (s.e. 0.39), confirming that instrumentation does not inflate residual inelasticity. The industry pass-through stays large and precise at 1.79 (s.e. 0.12), whereas the anomaly-style pass-through falls to 0.16 and is no longer statistically significant. Formally purging endogeneity therefore leaves the sector channel intact but removes most of the measured style spill-over. Table B.14 shows similar results when we use all-anomaly substitutes.

The IV results sharpen our earlier OLS message. Industry rotation is robust: it survives instrumentation and continues to absorb a substantial share of flow-induced price pressure. Style rebalancing, by contrast, cannot be distinguished from noise once its endogeneity is addressed—consistent with the idea that many anomaly portfolios still trade common stocks, so their returns contain considerable mechanical co-movement that the instrument strips away. Our nested-logit specification, which elevates industries to the outer nest while allowing cross-style heterogeneity through characteristic tastes in each inner nest, matches this empirical hierarchy: industry substitution is an important elastic margin, while style overlap contributes a modest adjustment.

B.3 Factor-Level Estimation

We also perform a fully factor-level demand estimation. Each quarter we build the investable set by double-sorting common stocks—first into ten Fama–French industries, then within every industry into quintiles on one stock characteristic at a time. The resulting industry–factor portfolios are treated as distinct "assets" in the Koijen–Yogo demand system. This construction internalises the substitution margin: investors can now rotate capital across factor tilts within a sector. Using ten rather than five industries enlarges each portfolio, ensuring that every institutional investor holds enough names to identify elasticities cleanly, but Section B.2.3 shows the substitutability patterns are the same under FF-5.

Following Li and Lin (2025), we focus on portfolios sorted by book-to-market. The trade-off is computational: running the full factor-level estimation and the associated variance decomposition on factor portfolios takes about 2.5 days per characteristic, even on our parallel cloud pipeline, which limits the exercise to the three core factors noted above. Despite the cost, this specification gives us the cleanest possible answer to the question: how our results change once factor rotation is built directly into the asset definition?

Construction of Factor Assets A direct "asset-level" implementation raises a practical hurdle: in the Koijen–Yogo framework an asset's identity must be fixed, yet factor portfolios are by construction re-sorted each quarter as firm characteristics drift. Stocks migrate in and out, so a portfolio defined at t is not the same object at $t + 1$; feeding such shifting baskets into the KY machinery would violate the model's premise that quantities and prices refer to the same asset across time. To preserve model consistency we therefore evaluate various portfolio sorts to design industry–factor portfolios that change only gradually. This approach keeps turnover low enough for the "asset" assumption to be credible, yet it still captures the cross-sectional tilts that motivate your proposed factor investing.

Table B.15 assesses how sticky portfolio membership is from one quarter to the next—a prerequisite for treating the industry–characteristic cells as quasi-fixed "assets" in the factor-level KY estimation. When stocks are sorted into book-to-market quintiles within each of the ten industries (Panel A) the average stay probability across sectors is 45%. Persistence is highest in the tails: the value-heavy Quintile 5 and the growth-heavy Quintile 1 both exceed 50% in every industry and reach 65% in Manufacturing and Utilities. Middle quintiles are less stable, hovering around 30–40%, but still well above pure chance. These figures imply that a large fraction of the book-to-market signal survives the quarterly re-balancing inherent in our sort.

Panel B shows that tightening the rank buckets to deciles reduces persistence, as expected, yet the decline is manageable. The cross-industry average stay probability drops to 28%, but the extreme deciles remain highly stable: Decile 1 and Decile 10 each post probabilities above 50% in most sectors. Because deciles also deliver twice as many observations per characteristic and therefore improve the precision of the demand estimates, we adopt the 10-industry×10-decile grid—100 basis assets—for the factor-level exercise. This choice maximises cross-sectional information while maintaining a durable asset definition that fits the KY framework.

Table B.16 repeats the persistence check using an "average-rank" value metric that pools book-to-market, sales-to-price, and related ratios before sorting. The retention figures drop noticeably: within-industry quintiles now keep an average stock for 35% of the next quarter (versus 45% for plain book-to-market), and deciles fall to 22% on average (versus 28%). Persistence remains acceptable in the extreme tails—Decile 1 and Decile 10 still exceed 35% in most industries—but the middle buckets turn over too rapidly. Because the higher-frequency reshuffling would blur the asset definition in our factor-level estimation, we retain book-to-market as the primary value sort.

Construction of Investor Factor Asset Holdings A further complication appears when we translate individual stock positions into our industry–value assets. Because each factor asset is defined by characteristics rather than by a fixed CUSIP list, two investors can "hold" the same asset without sharing a single constituent stock. Suppose the Manufacturing-Value factor contains 100 stocks. Investor 1 owns the first 50 and Investor 2 owns the remaining 50; together they cover the entire factor, yet their underlying universes are disjoint. If we credit each investor with a 50% position in the factor, the demand system treats them as trading the same asset even though their actual holdings never overlap, blurring price signals and exaggerating substitutability.

We recover each investor's exposure to the $F = 100$ industry–value assets with a data–driven projection rather than a mechanical stock–count rule. Following Van der Beck (2025), we approximate the investor's stock-weight vector $w_i(n)$ by a non-negative combination of the fully segmented industry-value decile portfolios $w^f(n)$. For quarter t we solve

$$\min_{\{\beta_t^f\}_{f=1}^F} \Big\| w_i(n) - \sum_{f=1}^{F} \beta_t^f \, w^f(n) \Big\|_2, \qquad 0 \le \beta_t^f \le 1 \ \ \forall f, \tag{B.15}$$

and, when desired, add the simplex constraint

$$\sum_{f=1}^{F} \beta_t^f = 1, \tag{B.16}$$

so that any residual automatically represents the investor's outside-asset demand. Because the demand system works with dollar weights, not share counts, the β_t^f coefficients serve directly as portfolio weights: the dollar allocation is $\beta_t^f \times \text{AUM}_{i,t}$.

To keep (B.15) well-conditioned we aggregate small accounts: investors with fewer than 200 distinct holdings are pooled by type until each projected matrix has at least 200 observations. These pooled entities control under 10% of total AUM, so the aggregation scarcely affects economic magnitudes.

The same projection framework will let us quantify how strongly each investor type tilts toward the value, investment, and profitability characteristics embedded in the industry–factor assets. By recording the R^2 from equation (B.15) for every investor and averaging it AUM-weightedly each quarter, we obtain a direct measure of how well systematic factor exposures explain real-world portfolios—no additional aggregation required. The numbers are encouraging: the industry–characteristic grid already explains roughly 74%–84% of the cross-sectional variation in holdings, with even higher fit for small investors, mutual funds, pension funds, and insurers.

With the fitted weights β_t^f, we estimate the Koijen et al. (2023) demand system on the $F = 100$ industry–value assets. We instrument each asset's price with its *level* counterfactual market equity, calculated from the holdings of all investors *except* the one whose demand we are measuring. A flow-based instrument—which works well in our reduced-form multiplier tests—would be less aligned with the KY framework here. KY relies on cross-sectional variation in the *levels* of investor positions to tackle the endogeneity of price levels. Flow shocks, by contrast, capture *changes* in positions and are therefore better suited to explain returns, not price levels. Hence we keep flow shocks for the reduced-form, return-based CFL tests in Section B.2, but use the level instrument for the cross-sectional KY estimation, where it provides a cleaner and more powerful identification.

We construct our factor-level instruments by aggregating the Koijen and Yogo (2019)

"$\widehat{\text{me}}$" variable up from individual stocks to entire portfolios. In the original KY setup, for each investor i and stock n, $\widehat{\text{me}}_i(n)$ captures exogenous demand for stock n coming from all other institutions.

To use this stock-level instrument in a portfolio-level regression, we compute a weighted average of $\widehat{\text{me}}_i(n)$ across all stocks n in a given portfolio. Specifically, if portfolio b holds a collection of stocks, we define $\widehat{\text{me}}_{i,t}(b)$ as the sum of $\widehat{\text{me}}_i(n)$ weighted by each stock's portfolio weight $w^b(n)$. In effect, $\widehat{\text{me}}_{i,t}(b)$ serves as the factor-level instrument for portfolio b, summarizing the exogenous variation in demand for all underlying stocks in that portfolio.

By aggregating $\widehat{\text{me}}_i(n)$ to $\widehat{\text{me}}_{i,t}(b)$, we capture a larger share of exogenous demand shocks—particularly those that arise when many institutions (e.g., large index funds) hold the same set of stocks. In KY's stock-level setting, the sparsity of institutional portfolios (with the median institution holding only 112 stocks) ensures that $\widehat{\text{me}}_i(n)$ is sufficiently idiosyncratic. At the portfolio level, however, we bundle more stocks together, increasing overlap across investors. On one hand, this pooling helps absorb exogenous demand shocks that affect widely held assets (for example, S&P500 constituents). On the other hand, fewer portfolios reduce cross-sectional heterogeneity, potentially weakening the instrument by diminishing variation in the investment universe across institutions. Using the same procedure, each factor-asset's characteristic vector is computed as the portfolio-weighted average of its constituent stocks' characteristics. These aggregate traits then feed directly into the regressions, ensuring that factor assets inherit the same information set as the underlying equities.

To verify that our aggregated instrument remains strong, we run first-stage regressions of log market equity (or log portfolio "price") on $\widehat{\text{me}}_{i,t}(b)$ and other controls, separately for each investor i at each date t. As shown in Table B.17, the AUM-weighted average first-stage t-statistic is 8.88, and 82 percent of institutions have t-stats above the Staiger–Stock (2005) critical value of 4.05. The cross-sectional distribution (mean 7.42, 1st percentile 3.21, 50th percentile 7.03, 99th percentile 14.84) confirms that only a small number of institutions fall below the weak-instrument cutoff. Hence, despite the reduced cross-section at the portfolio level, our instrument remains very strong for nearly all investors.

The two-step procedure—projection (B.15)–(B.16) followed by Koijen et al. (2023) estimation and variance decomposition—runs in roughly 2.5 days per characteristic sort on our parallel cloud setup, which confines the fully reported results to book-to-market, investment, and profitability while still delivering the clean factor-level test you requested.

Comparison of Stock-level and Portfolio-level Elasticities To engage directly with the substitution issue, we compare portfolio-level own-price elasticities under two demand specifications. First, we take the original stock-level KY estimates and aggregate them to the industry-factor portfolios; this preserves the simple-logit assumption of a single cross-elasticity and therefore omits substitution among the stocks inside each portfolio. Second, we re-estimate the demand system at the factor level, treating every industry-characteristic portfolio as its own asset so that within-portfolio substitution is built into the price coefficient. Both exercises rely on the same instruments and estimation mechanics, so any gap in elasticities reveals precisely how much extra price sensitivity emerges once the intra-portfolio substitute channel is allowed to operate.

In the stock–level specification each investor i has an estimated price–sensitivity coefficient $\beta_{0,i}$. The own–price elasticity for an individual stock n is

$$\varepsilon_i(n) \;=\; 1 - \beta_{0,i}\bigl(1 - w_i(n)\bigr),$$

where $w_i(n)$ is investor i's dollar weight in the stock. To compare these micro elasticities with factor–level results we collapse them to the same asset dimension. For every industry–characteristic portfolio f we compute a dollar–weighted average

$$\varepsilon_i^{\text{stock}\to f} \;=\; \sum_{n\in f} \frac{w_i(n)}{\sum_{m\in f} w_i(m)}\, \varepsilon_i(n),$$

yielding one elasticity per investor per factor asset.

In the factor specification the asset itself is portfolio f. The KY estimate $\beta_{0,i}^{f}$ enters utility once, giving the direct elasticity

$$\varepsilon_i^{\text{factor}}(f) \;=\; 1 - \beta_{0,i}^{f}\big(1 - \beta_i^{\text{alloc}}(f)\big),$$

where $\beta_i^{\text{alloc}}(f)$ is investor i's wealth share in asset f. Comparing $\varepsilon_i^{\text{stock}\to f}$ with $\varepsilon_i^{\text{factor}}(f)$ across all investors reveals how the portfolio elasticity estimates change when within–portfolio substitution is internalised in the asset definition.

The three elasticity tables (Table B.18–B.20) mirror the logic of our CFL reduced-form exercise but now in a structural setting. Panel A in each table takes the baseline stock-level model—which assumes a single cross-elasticity across all securities—and merely aggregates its elasticities up to the industry-factor portfolios. This calculation ignores the fact that, inside a given portfolio, investors can substitute one constituent stock for another. Panel B re-estimates the Koijen–Yogo system directly at the factor level, so each portfolio is itself an asset and all within-portfolio substitution is internalised. If the baseline specification is missing an important substitute channel, the factor-level elasticities should rise in exactly the same way the CFL multiplier fell once we controlled for the returns of close substitutes.

That is precisely what we see. For book-to-market portfolios (Table B.18) the baseline elasticities cluster around 0.30, while the factor-level estimates jump to about 0.50–0.60—

roughly a two-fold increase. Investment sorting in Table B.20 shows a similar pattern: stock-aggregated elasticities average 0.30 but climb to the mid-0.50s in the factor model. Profitability results in Table B.19 produces the largest gap; the aggregate elasticities again sit near 0.30, whereas the factor-level elasticities move up to 0.55–0.60, with several industries exceeding 0.70. Across all three characteristics, the inelasticity of portfolio demand is materially lower once we let investors reallocate freely among the stocks embedded in the factor asset.

The formal tests in Table B.21 come from the simple specification

$$\Delta\varepsilon_f \;=\; \alpha \;+\; u_f, \qquad \Delta\varepsilon_f \equiv \varepsilon_f^{\text{alt}} - \varepsilon_f^{\text{base}},$$

where f indexes every *industry* $\times$ *decile* factor portfolio and the dependent variable is the time–series mean elasticity difference for that portfolio. In Panel A the "alt" elasticity $\varepsilon_f^{\text{alt}}$ is estimated directly at the factor level using the composite-asset KY model, while $\varepsilon_f^{\text{base}}$ is the stock-level KY elasticity aggregated up to the same portfolio. Panel B reruns the test with $\varepsilon_f^{\text{alt}}$ taken from the *nested–logit* specification, which already embed within-industry substitution before the aggregation step. Because the regression contains only a constant, the coefficient α measures the average gap between the two elasticity estimates, and the reported t-statistic tests the null $H_0 : \alpha = 0$.

Across all three characteristics the factor–level specification generates elasticities roughly a quarter point above those implied by a straight aggregation of the simple-logit stock model—about 0.24 for value, 0.26 for profitability, and 0.25 for investment, all with t-statistics near eight. When we re-benchmark against the nested-logit stock model in which the outer nest is the coarser FF-5 industry grid narrows the average gap to about 0.13–0.14, yet it remains highly significant.

The smaller difference by about 0.1 is driven by an increase in elasticity when moving from the simple logit to the FF-5 nested-logit, and it is in line with the stock-level results in Table B.9. It indicates that substitution patterns captured at the stock level persist under aggregation, rather than washing out. In fact, the nested-logit framework closely aligns with the factor-level demand system, suggesting that accounting for industry-based substitution improves consistency across analysis at different level of aggregation. Overall, our findings are consistent with both Chaudhary et al. (2025) and Li and Lin (2025): measured elasticity tends to decline with aggregation, but the decline is less severe when heterogeneous substitution patterns are properly incorporated.

The combined results speak directly to the substitution concern. Under the simple–logit aggregation elasticities sit around 0.30 because the model forces a single cross-elasticity, masking investors' ability to shift capital among the stocks inside each industry-factor basket. Re-estimating the model on those baskets—so that this within-portfolio margin is taken into the account — raises elasticities to roughly 0.55, mirroring the CFL result that controlling for close substitutes reduces the price impact of flows. Aggregation still lowers elasticity overall, but once heterogeneous substitution is acknowledged—either by the nested logit or by direct factor-level estimation—the drop is smaller than in the simple logit.

Variance Decomposition Results Panel A of Table B.22 gives the main decomposition when book equity is included in the set of characteristics. Characteristics naturally dominate—the portfolios themselves are sorted on book-to-market, so the characteristic term is expected to be large—yet preferences remain prominent. On average, demand for stock characteristics accounts for 12% of the variance and latent demand contributes –4%. Those averages conceal substantial cross-industry dispersion. In Manufacturing, Other, and Telcom the demand-for-characteristics share climbs to 25–40%, while latent demand in Durables, Health, and Hi-Tech swings sharply negative. Large negative entries are not noise: they indicate that latent demand in those sectors tend to offset rather than amplify variation in the factor-based portfolios. Taken together, at least one of the two preference blocks (observable or latent) is sizable in every industry, confirming that investor taste continues to matter even when the value variable is built directly into both the asset definition and the demand estimation.

The 65% share labeled "Characteristics" warrants a brief clarification. Because the assets are literally formed on book-to-market deciles, book equity enters the variance twice—once by construction of the portfolio itself and again through its explicit appearance in investors' utility. This mechanical overlap is unavoidable and, if anything, gives the preference terms a tougher hurdle: they must compete with a characteristic that explains a very large slice of return movements on its own. The fact that the demand-for-characteristics block still explains double-digit variance shares in most industries underlines its economic relevance.

Panel B removes book equity from the preference specification while leaving the portfolio sort unchanged. As expected, the direct characteristic share collapses (65% to 8%) because the model no longer assigns a separate price component to that variable. Where does the variance go? Roughly one quarter migrates to demand-for-characteristics, which now rises to 23% on average, and latent demand creeps up to 5%. Industries such as Manufacturing, Hi-Tech, and Shops post demand-for-characteristics contributions in excess of 30%. The large residual term—64% versus 27% before—captures the portion of return variation formerly absorbed by book equity itself. What matters for our purposes is that the preference channels do not evaporate when the key observable is suppressed; if anything, they become more visible, signalling that investors still tilt toward value inside industries even when the model is "blind" to book equity.

The comparison across panels therefore strengthens rather than weakens our main conclusion. Investor preferences—both for observed characteristics and for motives that remain latent—retain substantial explanatory power whether we work with individual stocks or with industry-factor assets, and whether book equity is priced or omitted. Differences in the numerical shares reflect deliberate changes in what the model is allowed to label as "characteristic" versus "residual," not an erosion of the underlying economic forces.

Table B.1: Anomaly Portfolio HHI under FF-10 Industry Classification

This table reports the Herfindahl–Hirschman Indices (HHI) expressed as percentages for each anomaly portfolio's long and short legs under the Fama–French 10-industry classification. Panel A (long leg) and Panel B (short leg) show concentration measures across eight anomaly strategies. HHI is defined as $\sum_j w_j^2 \times 100$, where w_j is the portfolio weight in industry j.

Anomalies	Durbl	Enrgy	HiTec	Hlth	Manuf	NoDur	Other	Shops	Telcm	Utils	HHI
Panel A: Long Leg											
Mean	5%	8%	17%	9%	14%	8%	19%	12%	4%	5%	22%
Value	6%	8%	13%	7%	13%	6%	27%	7%	4%	8%	25%
Momentum	4%	8%	18%	9%	16%	7%	20%	10%	4%	4%	19%
Profitability	4%	7%	18%	11%	14%	10%	14%	14%	4%	3%	21%
Investment	4%	9%	17%	9%	17%	7%	15%	14%	4%	4%	18%
Issuance	4%	7%	18%	9%	11%	5%	27%	5%	7%	8%	19%
Size	5%	11%	15%	8%	15%	7%	18%	7%	7%	5%	14%
Tangibility	5%	6%	19%	7%	12%	9%	11%	27%	3%	1%	27%
Panel B: Short Leg											
Mean	4%	8%	15%	7%	14%	6%	27%	7%	6%	7%	23%
Value	4%	7%	18%	9%	15%	7%	21%	9%	5%	6%	21%
Momentum	4%	9%	17%	9%	15%	6%	20%	9%	6%	6%	20%
Profitability	4%	8%	10%	5%	11%	3%	36%	6%	7%	10%	28%
Investment	4%	7%	16%	8%	19%	8%	25%	6%	4%	3%	24%
Issuance	6%	15%	12%	6%	16%	9%	18%	7%	4%	5%	18%
Size	4%	7%	17%	10%	15%	8%	21%	10%	5%	4%	17%
Tangibility	4%	10%	20%	9%	13%	6%	19%	7%	10%	2%	22%

Table B.2: Anomaly Portfolio HHI under FF-5 Industry Classification

This table reports the Herfindahl–Hirschman Indices (HHI) expressed as percentages for each anomaly portfolio's long and short legs under the Fama–French five-industry classification. Panel A (long leg) and Panel B (short leg) show concentration measures across eight anomaly strategies. HHI is defined as $\sum_j w_j^2 \times 100$, where w_j is the portfolio weight in industry j.

Anomalies	Cnsmr	HiTec	Hlth	Manuf	Other	HHI
Panel A: Long Leg						
Mean	25%	21%	9%	26%	19%	31%
Value	20%	18%	7%	29%	27%	34%
Momentum	20%	22%	9%	28%	20%	28%
Profitability	28%	22%	11%	24%	14%	31%
Investment	25%	22%	9%	29%	15%	28%
Issuance	14%	25%	9%	26%	27%	27%
Size	19%	23%	8%	32%	18%	25%
Tangibility	41%	22%	7%	18%	11%	39%
Panel B: Short Leg						
Mean	17%	20%	8%	28%	27%	32%
Value	20%	23%	9%	28%	21%	30%
Momentum	19%	23%	9%	30%	20%	30%
Profitability	13%	17%	5%	29%	36%	35%
Investment	19%	20%	8%	29%	25%	32%
Issuance	23%	17%	6%	37%	18%	31%
Size	21%	22%	10%	25%	21%	26%
Tangibility	16%	30%	9%	26%	19%	32%

Table B.3: Observation Counts per Investor by Industry under FF-10 Classification

This table reports the mean (Panel A1 and B1) and minimum (Panel A2 and B2) number of stocks per investor across all sample quarters by investor type and industry under the Fama–French ten-industry classification. Panel A: investors with more than 1,000 holdings. Panel B: investors with ≤1,000 holdings.

Investor type	Durbl	Enrgy	HiTec	Hlth	Manuf	NoDur	Other	Shops	Telcm	Utils
Panel A: Investors with >1,000 holdings										
A1) Mean										
Small Investors	72	110	626	396	397	161	1048	289	70	93
Banks	47	75	366	217	264	102	555	200	46	78
Insurance Companies	46	74	350	205	263	101	562	201	45	80
Investment Advisors	36	63	288	163	210	82	425	162	37	67
Mutual Funds	47	76	372	223	264	101	575	203	47	74
Pension Funds	41	65	313	169	237	91	474	181	38	74
Other	38	68	294	174	219	86	454	169	40	68
Short Sellers	72	117	542	385	389	154	942	286	67	93
Insiders	50	79	436	310	276	89	683	199	33	63
Blockholders	63	95	597	376	364	146	799	300	60	39
A2) Minimum										
Small Investors	72	110	626	396	397	161	1048	289	70	93
Banks	23	42	182	87	146	61	276	106	24	49
Insurance Companies	29	51	219	109	175	70	333	129	26	64
Investment Advisors	19	32	163	76	127	48	249	99	19	27
Mutual Funds	23	38	187	90	143	55	274	108	17	28
Pension Funds	26	45	203	98	154	62	294	119	21	53
Other	30	53	230	117	170	67	329	131	30	51
Short Sellers	72	117	542	385	389	154	942	286	67	93
Insiders	50	79	436	310	276	89	683	199	33	63
Blockholders	63	95	597	376	364	146	799	300	60	39
Panel B: Investors with ≤1,000 holdings										
B1) Mean										
Banks	21	39	155	77	129	53	239	91	22	58
Insurance Companies	21	40	169	80	128	56	250	94	25	55
Investment Advisors	23	46	192	103	137	57	269	106	27	50
Mutual Funds	28	50	220	117	162	63	312	123	28	52
Pension Funds	24	44	182	84	142	59	265	105	26	57
Other	25	47	191	108	143	60	288	111	28	50
B2) Minimum										
Banks	12	25	89	45	79	35	156	57	15	44
Insurance Companies	16	32	129	59	102	46	191	73	20	44
Investment Advisors	11	26	111	55	82	35	171	62	16	26
Mutual Funds	16	36	162	77	116	44	214	84	19	32
Pension Funds	18	36	139	59	110	48	199	81	20	45
Other	16	35	144	74	107	47	210	83	21	36

Table B.4: Observation Counts per Investor by Industry under FF-5 Classification

This table reports the mean (Panel A1 and B1) and minimum (Panel A2 and B2) number of stocks per investor across all sample quarters by investor type and industry under the Fama–French five-industry classification. Panel A: investors with more than 1,000 holdings. Panel B: investors with ≤1,000 holdings.

Investor type	Cnsmr	HiTec	Hlth	Manuf	Other
Panel A: Investors with >1,000 holdings					
A1) Mean					
Small Investors	522	697	396	600	1048
Banks	350	412	217	417	555
Insurance Companies	347	394	205	417	562
Investment Advisors	281	325	163	340	425
Mutual Funds	351	418	223	414	575
Pension Funds	313	350	169	375	474
Other	292	334	174	355	454
Short Sellers	512	609	385	599	942
Insiders	337	468	310	418	683
Blockholders	509	658	376	498	799
A2) Minimum					
Small Investors	522	697	396	600	1048
Banks	193	208	87	249	276
Insurance Companies	230	247	109	293	333
Investment Advisors	176	189	76	206	249
Mutual Funds	193	210	90	235	274
Pension Funds	209	230	98	257	294
Other	229	263	117	283	329
Short Sellers	512	609	385	599	942
Insiders	337	468	310	418	683
Blockholders	509	658	376	498	799
Panel B: Investors with ≤1,000 holdings					
B1) Mean					
Banks	165	177	77	226	239
Insurance Companies	171	194	80	224	250
Investment Advisors	186	219	103	232	269
Mutual Funds	213	248	117	264	312
Pension Funds	189	208	84	243	265
Other	196	219	108	240	288
B2) Minimum					
Banks	107	105	45	155	156
Insurance Companies	137	151	59	184	191
Investment Advisors	115	132	55	149	171
Mutual Funds	151	185	77	199	214
Pension Funds	149	161	59	196	199
Other	149	168	74	189	210

Table B.5: Average Portfolio Weights in Outside Assets by Industry and Investor Type

This table reports the average portfolio weights (in percentages) allocated to outside assets for each industry across investor types. Panel A uses the KY definition of outside assets, while Panel B uses the KY+KRY definition.

Investor type	Cnsmr	HiTec	Hlth	Manuf	Other
Panel A: KY definition					
Small Investors	2.71%	4.74%	1.48%	4.56%	10.43%
Banks	0.96%	1.62%	0.39%	1.99%	4.02%
Insurance Companies	1.03%	2.40%	0.49%	2.36%	8.19%
Investment Advisors	1.67%	3.75%	1.14%	2.93%	7.78%
Mutual Funds	1.35%	2.94%	0.84%	2.53%	6.56%
Pension Funds	1.25%	3.14%	0.53%	2.78%	8.94%
Other	2.48%	6.01%	2.07%	3.45%	10.16%
Short Sellers	1.76%	3.61%	1.11%	2.46%	6.69%
Insiders	0.94%	1.72%	0.75%	1.65%	6.12%
Blockholders	1.77%	5.15%	0.90%	1.46%	6.25%
Panel B: KY+KRY definition					
Small Investors	0.20%	0.27%	0.14%	0.29%	0.65%
Banks	0.07%	0.05%	0.01%	0.09%	0.44%
Insurance Companies	0.15%	0.22%	0.03%	0.22%	0.51%
Investment Advisors	0.46%	0.58%	0.21%	0.67%	1.09%
Mutual Funds	0.23%	0.34%	0.11%	0.37%	0.36%
Pension Funds	0.04%	0.12%	0.01%	0.08%	0.11%
Other	0.73%	0.92%	0.34%	1.09%	1.57%
Short Sellers	0.19%	0.18%	0.06%	0.30%	0.20%
Insiders	0.47%	0.56%	0.22%	0.68%	1.47%
Blockholders	0.35%	0.42%	0.17%	0.45%	0.53%

Table B.6: Unique Outside-Asset Counts and Percentiles by Industry and Investor Type

This table reports the average number of unique outside assets (Panel A) and the cross-sectional percentiles (Panels B–E) by industry and investor type. Percentiles correspond to the 1st, 5th, 10th, and 20th percentiles of the asset-count distribution.

Investor type	Cnsmr	HiTec	Hlth	Manuf	Other
Panel A: Mean					
Banks	9	17	7	13	44
Insurance Companies	9	17	8	12	50
Investment Advisors	3	5	2	4	13
Mutual Funds	9	18	8	12	45
Pension Funds	12	21	8	15	64
Other	3	6	2	5	15
Panel B: 1st percentile					
Banks	0	0	0	0	0
Insurance Companies	0	0	0	0	0
Investment Advisors	0	0	0	0	0
Mutual Funds	0	0	0	0	0
Pension Funds	0	0	0	0	0
Other	0	0	0	0	0
Panel C: 5th percentile					
Banks	0.1	0.1	0.0	0.2	0.6
Insurance Companies	0.0	0.0	0.0	0.1	0.4
Investment Advisors	0.0	0.0	0.0	0.0	0.0
Mutual Funds	0.0	0.1	0.0	0.0	0.6
Pension Funds	0.1	0.2	0.0	0.4	1.1
Other	0.0	0.0	0.0	0.0	0.0
Panel D: 10th percentile					
Banks	0.3	0.4	0.1	0.8	1.1
Insurance Companies	0.0	0.2	0.0	0.2	1.0
Investment Advisors	0.0	0.0	0.0	0.0	0.3
Mutual Funds	0.1	0.6	0.0	0.5	1.6
Pension Funds	0.4	0.8	0.0	0.8	2.2
Other	0.0	0.0	0.0	0.0	0.1
Panel E: 20th percentile					
Banks	0.7	0.8	0.2	1.5	2.2
Insurance Companies	0.4	1.0	0.1	0.9	2.0
Investment Advisors	0.0	0.3	0.0	0.1	1.1
Mutual Funds	0.7	1.5	0.2	1.4	3.6
Pension Funds	1.3	2.3	0.4	1.9	6.8
Other	0.0	0.2	0.0	0.0	0.9

Table B.7: Fraction of Investor AUM Reclassified as Small Investor under Outside-Asset Cutoff Thresholds

This table reports the average fraction of assets under management (in percentages) reclassified as household outside assets for each investor type under varying cutoff thresholds $x\%$. Outside assets are defined as the bottom $x\%$ of stocks by industry (FF-5) plus KR's outside-asset definition.

Cutoff ($x\%$)	Banks	Insurance Companies	Investment Advisors	Mutual Funds	Pension Funds	Other
1%	1.19%	6.61%	21.99%	5.17%	3.95%	17.14%
2%	1.19%	6.61%	21.97%	5.16%	3.95%	17.13%
3%	1.19%	6.61%	21.93%	5.15%	3.95%	16.96%
4%	1.19%	6.61%	21.88%	5.14%	3.95%	16.84%
5%	1.18%	6.61%	21.85%	5.14%	3.95%	16.71%
6%	1.18%	6.61%	21.81%	5.13%	3.95%	16.69%
7%	1.18%	6.60%	21.77%	5.11%	3.95%	16.28%
8%	1.17%	6.60%	21.73%	5.10%	3.95%	16.27%
9%	1.17%	6.59%	21.68%	5.10%	3.95%	16.22%
10%	1.16%	6.49%	21.63%	5.09%	3.95%	16.20%
11%	1.16%	6.49%	21.56%	5.09%	3.92%	16.14%
12%	1.15%	6.48%	21.46%	5.07%	3.92%	16.11%
13%	1.15%	6.48%	21.41%	5.06%	3.90%	16.05%
14%	1.15%	6.37%	21.31%	5.03%	3.90%	15.97%
15%	1.14%	6.36%	21.25%	5.02%	3.87%	15.85%

Table B.8: Substitution Elasticities by Industry and Investor Group

This table reports time-series means of cross-sectional substitution elasticities σ for each industry and investor type. Columns represent investor groups; rows represent industries.

Industry	Small Investors	Banks	Insurance Comp.	Invest. Advisors	Mutual Funds	Pension Funds	Other	Short Sellers	Blockholders
Cnsmr	0.81	0.27	0.47	0.37	0.36	0.37	0.38	0.21	0.40
Hlth	1.00	0.40	0.63	0.48	0.52	0.55	0.51	0.01	0.57
HiTec	0.88	0.20	0.28	0.24	0.22	0.23	0.26	0.00	0.00
Manuf	0.76	0.23	0.38	0.31	0.34	0.30	0.30	0.31	0.58
Other	0.47	0.15	0.19	0.18	0.20	0.20	0.12	0.18	0.25

Table B.9: Comparison of Stock-Level Elasticities Across Investors and Demand Models

Panel A reports means and standard deviations for the simple logit model across investor types. Panel B reports analogous statistics for the nested logit model. Panel C reports coefficient differences (Nested – Logit) and t-statistics from equal-weight tests.

	Small Investors	Banks	Insurance Companies	Investment Advisors	Mutual Funds	Pension Funds	Other	Short Sellers	Blockholders
Panel A: Simple Logit									
Mean	0.65	0.08	0.13	0.42	0.15	0.10	0.33	0.01	0.09
Std	0.00	0.13	0.13	0.21	0.12	0.10	0.25	0.00	0.00
Panel B: Nested Logit									
Mean	0.62	0.14	0.21	0.54	0.17	0.17	0.39	0.03	0.13
Std	0.15	0.24	0.26	0.40	0.18	0.20	0.45	0.05	0.10
Panel C: Difference (Nested – Simple Logit) in Investor Elasticity									
Differences	−0.02	0.06	0.09	0.12	0.02	0.07	0.06	0.02	0.04
t-stat	−1.64	11.89	11.75	13.37	5.15	15.53	5.92	3.26	12.06

Table B.10: Comparison of Portfolio Elasticities Across Demand Models

(a) Panel A: Elasticities by Anomaly Portfolio (Long vs. Short)

	Simple Logit		Nested Logit	
Anomaly	Long Leg	Short Leg	Long Leg	Short Leg
Momentum	0.28	0.28	0.38	0.40
Investment	0.30	0.29	0.40	0.39
Profitability	0.30	0.30	0.40	0.40
Tangibility	0.29	0.28	0.39	0.38
Value	0.31	0.29	0.41	0.38
Size	0.33	0.32	0.43	0.40
Issuance	0.31	0.28	0.41	0.38

This panel reports, for each anomaly portfolio, the long-leg and short-leg elasticity estimates under the simple logit model and the nested logit model.

(b) Panel B: Difference (Nested – Simple Logit) in Elasticities

	Short Leg	Q2	Q3	Q4	Long Leg	Pooled
Difference	0.09	0.10	0.10	0.10	0.10	0.10
t-stat	10.97	11.22	11.10	10.78	10.42	10.42

This panel reports differences in mean elasticities (nested – simple logit) for the short-leg, intermediate anomaly portfolio quintiles (2, 3, 4), the long-leg, and the pooled sample, along with corresponding t-statistics.

Table B.11: CFL Stock-Level Flow-Multiplier Estimates under Different FE Specifications

This table reports the estimated multiplier (coefficient on the own shock $\mathcal{F}_{i,t}$) from CFL stock-level regressions under four fixed-effect specifications. Standard errors in parentheses.

	(1)	(2)	(3)	(4)
Own shock $\mathcal{F}_{i,t}$	0.175***	0.158***	0.153***	0.153***
	(0.0350)	(0.0324)	(0.0313)	(0.0313)
Observations	225,681	225,681	225,681	225,681
R-squared	0.216	0.242	0.258	0.258
Time FE	Yes	—	—	—
FF5 × Time FE	No	Yes	—	—
FF10 × Time FE	No	No	Yes	—
FF49 × Time FE	No	No	No	Yes

Standard errors in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table B.12: CFL Portfolio-Level Regressions under Different Industry Definitions

This table reports the estimated coefficients from portfolio-level regressions of returns on the own shock $F_{P,t}$ and industry control $R_{P,t}^{\text{Ind}}$ under three industry grids: FF5, FF10, and FF49. Standard errors in parentheses.

	(1)	(2)	(3)	(4)
Own shock $F_{P,t}$	2.536***	1.559***	1.374***	1.165***
	(0.597)	(0.387)	(0.372)	(0.330)
Industry substitute return $R_{P,t}^{\text{Ind}}$	—	1.798***	1.557***	1.379***
	—	(0.0743)	(0.0668)	(0.0542)
Quarter FE	Yes	Yes	Yes	Yes
Industry Grid		FF5	FF10	FF49
Observations	6,532	6,532	6,532	6,532
R-squared	0.887	0.930	0.940	0.954

Standard errors in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table B.13: CFL Portfolio-Level Regressions with Close Substitutes

This table reports portfolio-level regression coefficients of returns on own shocks $F_{P,t}$, substitute returns $R^{\text{sub}}_{P,t}$, anomaly substitute shocks $F^{\text{sub}}_{P,t}$, industry controls $R^{\text{Ind}}_{P,t}$, and industry shocks $F^{\text{Ind}}_{P,t}$ under various FE and IV specifications.

	(1)	(2)	(3)	(4)	(5) FS_{sub}	(6) FS_{Ind}	(7) 2SLS
Dependent var.	$R_{P,t}$	$R_{P,t}$	$R_{P,t}$	$R_{P,t}$	$R_{\text{sub},P,t}$	$R^{\text{Ind}}_{P,t}$	$R_{P,t}$
Own shock $F_{P,t}$	2.536***	1.559***	1.575***	1.554***			1.530***
	(0.597)	(0.387)	(0.399)	(0.393)			(0.385)
Anomaly substitute return $R_{\text{sub},P,t}$				0.183**			0.156
				(0.0583)			(0.091)
Anomaly substitute shock $F^{\text{sub}}_{P,t}$					5.459***	1.347**	
					(1.446)	(0.457)	
Industry substitute return $R^{\text{Ind}}_{P,t}$		1.798***	1.814***	1.736***			1.794***
		(0.0743)	(0.0757)	(0.0694)			(0.115)
Industry substitute shock $F^{\text{Ind}}_{P,t}$					0.458***	1.128***	
					(0.066)	(0.104)	
Cragg–Donald F-statistic					234.57	234.57	
Category × Quarter FE	No	No	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	No	No	No	No	No
Observations	6,532	6,532	6,532	6,532	6,532	6,532	6,532
R-squared	0.887	0.930	0.936	0.937	0.970	0.990	0.937

Standard errors in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table B.14: CFL Portfolio-Level Regressions: All Anomaly Portfolios as Substitutes

This table reports portfolio-level regression coefficients of returns on own shocks $F_{P,t}$, substitute returns $R_{P,t}^{\text{sub}}$, anomaly substitute shocks $F_{P,t}^{\text{sub}}$, industry controls $R_{P,t}^{\text{Ind}}$, and industry shocks $F_{P,t}^{\text{Ind}}$ under various FE and IV specifications.

	(1)	(2)	(3)	(4)	(5) FS_{sub}	(6) FS_{Ind}	(7) 2SLS
Dependent var.	$R_{P,t}$	$R_{P,t}$	$R_{P,t}$	$R_{P,t}$	$R_{\text{sub},P,t}$	$R_{P,t}^{\text{Ind}}$	$R_{P,t}$
Own shock $F_{P,t}$	2.536***	1.559***	1.575***	1.541***			1.532***
	(0.597)	(0.387)	(0.399)	(0.390)			(0.384)
Substitute return $R_{\text{sub},P,t}$				0.540***			0.246
				(0.051)			(0.145)
Anomaly substitute shock $F_{P,t}^{\text{sub}}$					3.449***	1.347**	
					(0.899)	(0.457)	
Industry control $R_{P,t}^{\text{Ind}}$		1.798***	1.814***	1.667***			1.794***
		(0.074)	(0.076)	(0.066)			(0.113)
Industry substitute shock $F_{P,t}^{\text{Ind}}$					0.287***	1.128***	
					(0.041)	(0.104)	
Cragg–Donald F-statistic					401.49	401.49	
Category × Quarter FE	No	No	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	No	No	No	No	No
Observations	6,532	6,532	6,532	6,532	6,532	6,532	6,532
R-squared	0.887	0.930	0.936	0.938	0.992	0.990	0.937

Standard errors in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table B.15: Sorting on B/M: Probability that the Stock Remains in the Same Portfolio Next Quarter

Panel A reports, for each industry, the probability (in percentages) that a stock assigned to a given book-to-market quintile remains in that same quintile in the following quarter. Panel B reports analogous probabilities when stocks are sorted into deciles by book-to-market within industry. The last row of each panel shows the cross-industry average.

	Durbl	Enrgy	HiTec	Hlth	Manuf	NoDur	Other	Shops	Telcm	Utils
Panel A: Quintiles										
Quintile 1	58%	52%	54%	54%	64%	67%	53%	62%	52%	65%
Quintile 2	39%	29%	36%	27%	46%	50%	43%	41%	31%	45%
Quintile 3	34%	25%	31%	27%	38%	41%	40%	37%	34%	39%
Quintile 4	32%	30%	32%	28%	38%	44%	45%	36%	34%	42%
Quintile 5	64%	52%	60%	56%	63%	60%	60%	59%	59%	61%
Average	46%	38%	42%	38%	50%	53%	48%	47%	42%	50%
Panel B: Deciles										
Decile 1	52%	47%	47%	46%	54%	53%	48%	55%	45%	54%
Decile 2	27%	19%	28%	19%	39%	40%	32%	36%	23%	30%
Decile 3	25%	18%	20%	16%	29%	34%	27%	25%	19%	28%
Decile 4	21%	17%	18%	12%	24%	25%	24%	25%	21%	18%
Decile 5	17%	14%	18%	11%	19%	25%	23%	19%	18%	24%
Decile 6	19%	13%	17%	16%	22%	23%	22%	20%	20%	20%
Decile 7	17%	15%	17%	14%	20%	27%	24%	20%	18%	25%
Decile 8	18%	17%	17%	17%	23%	30%	28%	18%	19%	20%
Decile 9	20%	24%	25%	22%	26%	29%	33%	27%	29%	31%
Decile 10	56%	40%	53%	48%	57%	52%	53%	50%	54%	59%
Average	27%	22%	26%	22%	31%	34%	31%	29%	27%	31%

Table B.16: Sorting on Averaged Ranked Value: Probability that the Stock Remains in the Same Portfolio Next Quarter

Panel A reports, for each industry, the probability (in percentages) that a stock assigned to a given quintile based on averaged rank of value characteristics remains in that same quintile in the following quarter. Panel B reports analogous probabilities when stocks are sorted into deciles by averaged rank of value. The last row of each panel shows the cross-industry average.

	Durbl	Enrgy	HiTec	Hlth	Manuf	NoDur	Other	Shops	Telcm	Utils
Panel A: Quintiles										
Quintile 1	47%	37%	51%	46%	50%	54%	57%	56%	41%	52%
Quintile 2	9%	21%	18%	22%	11%	21%	0%	11%	8%	27%
Quintile 3	54%	15%	0%	16%	43%	22%	0%	50%	38%	29%
Quintile 4	28%	26%	51%	46%	43%	32%	0%	33%	31%	28%
Quintile 5	47%	52%	48%	54%	46%	46%	77%	46%	49%	46%
Average	37%	30%	34%	37%	39%	35%	27%	39%	33%	36%
Panel B: Deciles										
Decile 1	43%	32%	42%	39%	42%	46%	48%	51%	39%	47%
Decile 2	17%	13%	24%	16%	22%	22%	22%	21%	16%	16%
Decile 3	9%	17%	15%	12%	9%	12%	0%	10%	5%	16%
Decile 4	0%	10%	5%	15%	7%	7%	0%	2%	21%	15%
Decile 5	27%	5%	0%	8%	0%	11%	0%	25%	23%	22%
Decile 6	25%	7%	0%	18%	40%	15%	0%	23%	18%	14%
Decile 7	19%	8%	31%	39%	31%	18%	0%	17%	13%	16%
Decile 8	12%	16%	32%	12%	17%	18%	0%	16%	17%	16%
Decile 9	20%	27%	20%	26%	19%	17%	73%	21%	14%	18%
Decile 10	36%	35%	37%	42%	38%	40%	47%	39%	40%	34%
Average	21%	17%	21%	23%	23%	21%	19%	23%	20%	22%

Table B.17: Distribution of t-Statistics Across Institutions

Panel A: AUM-Weighted t and Exceedance Rate	
AUM-weighted $\bar{t}_{\mathrm{IV}}$	% t-stats > 4.05
8.88	0.82

Panel B: Cross-Sectional Distribution of t-Statistics

Mean	Std	Min	1%	5%	10%	25%	50%	75%	90%	95%	99%	Max
7.42	2.40	1.47	3.21	4.22	4.87	5.82	7.03	8.50	10.61	12.38	14.84	15.41

Table B.18: Elasticity Estimates for B/M-Industry Portfolios

This table reports time-series means of elasticity estimates for book-to-market industry portfolios. Panel A shows bottom-up (LST) stock-level elasticities aggregated to the portfolio level. Panel B shows factor-level (composite-asset) elasticities. Each column corresponds to a B/M rank from Growth (lowest) to Value (highest). The final row of each panel is the cross-industry average.

	Growth	2	3	4	5	6	7	8	9	Value
Panel A: Stock-Level Aggregated										
Durbl	0.20	0.22	0.28	0.28	0.31	0.25	0.26	0.33	0.31	0.31
Enrgy	0.29	0.31	0.28	0.32	0.31	0.30	0.28	0.27	0.26	0.30
HiTec	0.28	0.29	0.30	0.32	0.30	0.27	0.29	0.30	0.34	0.38
Hlth	0.27	0.30	0.30	0.31	0.34	0.28	0.30	0.25	0.27	0.30
Manuf	0.30	0.29	0.28	0.30	0.30	0.26	0.31	0.27	0.26	0.35
NoDur	0.37	0.34	0.32	0.31	0.32	0.32	0.31	0.33	0.29	0.32
Other	0.29	0.28	0.28	0.29	0.31	0.32	0.32	0.33	0.32	0.32
Shops	0.31	0.28	0.31	0.30	0.31	0.27	0.26	0.22	0.24	0.29
Telcm	0.34	0.30	0.32	0.35	0.33	0.30	0.29	0.29	0.37	0.39
Utils	0.31	0.33	0.44	0.36	0.35	0.32	0.33	0.34	0.33	0.29
Average	0.30	0.29	0.31	0.31	0.32	0.29	0.30	0.29	0.30	0.32
Panel B: Factor Level										
Durbl	0.61	0.59	0.57	0.60	0.61	0.60	0.62	0.63	0.62	0.58
Enrgy	0.56	0.48	0.58	0.55	0.51	0.56	0.58	0.55	0.56	0.65
HiTec	0.43	0.43	0.47	0.48	0.50	0.49	0.58	0.55	0.62	0.69
Hlth	0.39	0.43	0.41	0.49	0.55	0.56	0.50	0.65	0.61	0.67
Manuf	0.41	0.43	0.43	0.49	0.46	0.61	0.62	0.56	0.61	0.80
NoDur	0.42	0.43	0.51	0.54	0.57	0.56	0.63	0.58	0.63	0.59
Other	0.50	0.45	0.46	0.44	0.44	0.43	0.45	0.50	0.52	0.69
Shops	0.43	0.45	0.49	0.56	0.60	0.60	0.61	0.66	0.69	0.75
Telcm	0.68	0.58	0.55	0.47	0.47	0.47	0.54	0.56	0.63	0.61
Utils	0.52	0.51	0.62	0.53	0.48	0.56	0.51	0.54	0.57	0.55
Average	0.50	0.48	0.51	0.52	0.52	0.54	0.56	0.58	0.61	0.66

Table B.19: Elasticity Estimates for Profitability-Industry Portfolios

This table reports time-series means of elasticity estimates for profitability-industry portfolios. Panel A shows bottom-up stock-level elasticities aggregated to the portfolio level. Panel B shows factor-level elasticities. Each column corresponds to an profitability rank from Weak (lowest) to Robust (highest). The final row of each panel is the cross-industry average.

	Weak	2	3	4	5	6	7	8	9	Robust
Panel A: Stock-Level Aggregated										
Durbl	0.30	0.29	0.30	0.28	0.32	0.27	0.24	0.22	0.22	0.21
Enrgy	0.34	0.25	0.28	0.26	0.32	0.36	0.30	0.30	0.26	0.35
HiTec	0.32	0.28	0.29	0.31	0.30	0.30	0.29	0.29	0.29	0.33
Hlth	0.28	0.27	0.27	0.26	0.27	0.32	0.28	0.29	0.32	0.29
Manuf	0.27	0.28	0.27	0.29	0.30	0.26	0.30	0.30	0.30	0.29
NoDur	0.35	0.31	0.31	0.34	0.34	0.33	0.28	0.34	0.35	0.32
Other	0.35	0.36	0.30	0.30	0.31	0.31	0.31	0.28	0.28	0.26
Shops	0.26	0.24	0.27	0.24	0.23	0.30	0.32	0.38	0.27	0.27
Telcm	0.31	0.32	0.33	0.32	0.34	0.35	0.31	0.33	0.39	0.33
Utils	0.39	0.35	0.32	0.31	0.35	0.36	0.34	0.36	0.30	0.29
Average	0.32	0.29	0.29	0.29	0.31	0.32	0.30	0.31	0.30	0.29
Panel B: Factor Level										
Durbl	0.58	0.60	0.58	0.63	0.58	0.64	0.57	0.61	0.69	0.60
Enrgy	0.70	0.55	0.50	0.58	0.61	0.63	0.54	0.61	0.60	0.66
HiTec	0.66	0.64	0.49	0.57	0.55	0.54	0.50	0.49	0.52	0.57
Hlth	0.58	0.51	0.59	0.44	0.56	0.58	0.43	0.43	0.47	0.50
Manuf	0.53	0.55	0.50	0.50	0.59	0.48	0.52	0.50	0.51	0.49
NoDur	0.58	0.53	0.62	0.61	0.59	0.58	0.50	0.54	0.57	0.53
Other	0.56	0.55	0.47	0.51	0.52	0.49	0.50	0.51	0.51	0.50
Shops	0.54	0.50	0.61	0.52	0.52	0.70	0.52	0.65	0.54	0.56
Telcm	0.61	0.49	0.55	0.63	0.61	0.57	0.53	0.57	0.53	0.56
Utils	0.73	0.61	0.52	0.56	0.51	0.66	0.57	0.64	0.56	0.59
Average	0.61	0.55	0.54	0.56	0.56	0.59	0.52	0.56	0.55	0.56

Table B.20: Elasticity Estimates for Investment-Industry Portfolios

This table reports time-series means of elasticity estimates for investment-industry portfolios when sorted from Aggressive to Conservative. Panel A shows bottom-up (LST) stock-level elasticities aggregated to the portfolio level. Panel B shows factor-level (composite-asset) elasticities. Each column corresponds to an investment rank, from Aggressive (lowest) to Conservative (highest). The final row of each panel is the cross-industry average.

	Aggress.	2	3	4	5	6	7	8	9	Conservative
Panel A: Stock-Level Aggregated										
Durbl	0.22	0.29	0.33	0.31	0.27	0.26	0.20	0.24	0.23	0.32
Enrgy	0.34	0.34	0.25	0.28	0.27	0.32	0.29	0.31	0.32	0.28
HiTec	0.26	0.27	0.31	0.28	0.29	0.31	0.29	0.32	0.31	0.34
Hlth	0.27	0.26	0.27	0.29	0.31	0.33	0.31	0.28	0.29	0.30
Manuf	0.27	0.26	0.28	0.28	0.29	0.29	0.29	0.30	0.30	0.32
NoDur	0.31	0.28	0.32	0.33	0.31	0.34	0.33	0.35	0.33	0.33
Other	0.27	0.30	0.30	0.30	0.32	0.31	0.33	0.31	0.30	0.31
Shops	0.28	0.24	0.27	0.28	0.31	0.30	0.28	0.27	0.24	0.24
Telcm	0.30	0.30	0.32	0.34	0.32	0.35	0.32	0.30	0.28	0.30
Utils	0.30	0.35	0.38	0.34	0.34	0.34	0.32	0.33	0.33	0.31
Average	0.28	0.29	0.30	0.30	0.30	0.32	0.30	0.30	0.29	0.31
Panel B: Factor Level										
Durbl	0.65	0.60	0.61	0.66	0.54	0.56	0.58	0.63	0.64	0.61
Enrgy	0.66	0.72	0.56	0.60	0.58	0.60	0.57	0.54	0.55	0.63
HiTec	0.48	0.50	0.52	0.46	0.52	0.51	0.47	0.51	0.58	0.59
Hlth	0.50	0.50	0.48	0.57	0.47	0.55	0.50	0.48	0.51	0.54
Manuf	0.48	0.50	0.48	0.45	0.50	0.49	0.49	0.51	0.63	0.64
NoDur	0.66	0.50	0.49	0.55	0.53	0.58	0.55	0.57	0.58	0.64
Other	0.50	0.55	0.45	0.47	0.51	0.50	0.50	0.48	0.59	0.58
Shops	0.57	0.48	0.46	0.47	0.59	0.65	0.56	0.53	0.58	0.62
Telcm	0.60	0.59	0.59	0.56	0.59	0.57	0.62	0.45	0.61	0.59
Utils	0.61	0.60	0.66	0.54	0.54	0.54	0.60	0.55	0.60	0.58
Average	0.57	0.55	0.53	0.53	0.54	0.55	0.54	0.53	0.59	0.60

Table B.21: Regression of Elasticity Differences on a Constant

Panel A reports results using factor-level elasticity estimates. Panel B reports results using nested-logit elasticity estimates. Each regression is of the difference between factor-level and stock-level elasticities on a constant.

	Difference	
Asset	Coefficient	(t-stat)
Panel A: Factor-Level Estimates		
Value	0.24	(8.63)
Profitability	0.26	(8.00)
Investment	0.25	(7.68)
Panel B: Nested-Logit Estimates		
Value	0.13	(4.37)
Profitability	0.14	(4.31)
Investment	0.14	(3.91)

Table B.22: Variance Decomposition in B/M-Industry Portfolios: With and Without Book Equity

Panel A reports, for each industry, the proportion of total variance attributed to Demand for Stock Characteristics, Latent Demand, Characteristics, and Residual Effects when book equity is included. Panel B reports the analogous proportions when book equity is excluded. Each panel concludes with the cross-industry average.

Industry	Demand for Stock Characteristics	Latent Demand	Characteristics	Residual Effects
Panel A: Estimation with Book Equity				
Durbl	24%	−17%	52%	41%
Enrgy	7%	−7%	79%	21%
HiTec	−4%	−26%	85%	44%
Hlth	−19%	14%	58%	46%
Manuf	25%	6%	58%	11%
NoDur	4%	10%	47%	38%
Other	40%	−12%	69%	4%
Shops	17%	1%	29%	51%
Telcm	27%	−9%	65%	17%
Utils	−6%	2%	112%	−7%
Average	12%	−4%	65%	27%
Panel B: Estimation without Book Equity				
Durbl	3%	−11%	26%	83%
Enrgy	27%	33%	−1%	40%
HiTec	36%	−72%	56%	83%
Hlth	3%	24%	6%	66%
Manuf	88%	−2%	−22%	33%
NoDur	4%	29%	−6%	74%
Other	13%	17%	11%	60%
Shops	45%	−9%	23%	40%
Telcm	−41%	31%	−1%	110%
Utils	48%	11%	−8%	50%
Average	23%	5%	8%	64%

C Additional Results

C.1 Are the Demand Effects of Small Investors Mechanical?

C.1.1 Dollar AUM Invested in Anomaly Portfolios by Investor Groups

We examine two mechanical explanations for the effects of small investors. One possibility is that the substantial effects of their demand are driven by their larger holdings in anomaly portfolios.[28] To investigate this, we calculate the fraction of the total dollar value of anomaly portfolios held by different investor types using the following approach. First, for each stock n, we calculate the total dollar position $h_t^{(g)}(n)$ held by investor group g as:

$$h_t^{(g)}(n) = \sum_{i \in g} w_{i,t}(n) AUM_{i,t}. \tag{C.1}$$

Then, for each individual anomaly p at time t, we sum $h_t^{(g)}(n)$ over all the stocks that belong to leg l of the portfolio:

$$h_{p,t}^{l,(g)} = \sum_{n \in l} h_t^{(g)}(n). \tag{C.2}$$

We then normalize $h_{p,t}^{l,(g)}$ by dividing it by the total dollar value of all the stocks in leg l of anomaly p. Averaging across all individual anomalies within the same anomaly group, we report the sample means.

Appendix Table C.1 reveals that, contrary to this hypothesis, small investors held less than 31% of the dollar value of the long leg and less than 32% of the short leg for each anomaly group from 2006-2022. The only exception is the early 1999–2005 period, when

[28] For instance, if small investors hold around 50% of the momentum portfolio's combined dollar value, it is unsurprising that their total demand explains 60% of the variation in momentum returns. The influence of any investor's demand on asset prices is stronger when that investor holds a significant fraction of the asset.

small investors hold about 40%, but our results are not driven by these initial years, and even this share is small relative to the fraction of anomaly variance their effects explain. This finding underscores the significant impact of small investors, suggesting their influence is disproportionately large relative to the dollar size of their holdings.

C.1.2 Percentage of Stocks in Anomaly Portfolios Held by Investor Groups

Another possibility is that small investors and institutions do not hold the same stocks within anomaly portfolios. To explore this, we calculate the fraction of stocks from each anomaly portfolio held by each investor type using the following procedure. For each individual anomaly p at time t, we compute the fraction of stocks with positive positions within each investor group's investment universe, separately for the long and short legs of the anomaly portfolio. For leg l, this fraction is defined as:

$$IS_{p,t}^{l,(g)}\% = \frac{\textit{Number of stocks by investor group } g \textit{ in leg } l \textit{ for anomaly } p \textit{ at time } t}{\textit{Total number of stocks in leg } l \textit{ for anomaly } p \textit{ at time } t}. \quad \text{(C.3)}$$

We then average $IS_{p,t}^{l,(g)}$ across all individual anomalies within the same anomaly group and report the sample means.

Appendix Table C.2 shows that small investors hold only a marginally higher fraction of stocks. For example, small investors hold 97% of positions in the long leg of the average anomaly portfolio, while banks, mutual funds, and investment advisors hold 89%, 88%, and 86%, respectively. This small difference is expected because our definition of small investors relies on residual ownership; stocks not held by institutions are attributed to small investors. However, this difference is unlikely to explain the much larger effect of small investors' demand, especially given their smaller dollar holdings in the anomaly portfolios.[29]

[29]The results in Appendix Table C.2 use the full universe of U.S. stocks for completeness, while our main

Our combined findings indicate that the disproportionate influence of small investors cannot be attributed solely to their portfolio size or stock holdings within anomaly portfolios. The persistent and significant impact of their demand on anomaly returns suggests that small investors exhibit unique trading patterns that distinguish their investment strategies from those of institutional investors.

C.2 Small Investors: Validation using Cross Sectional Tests

Our "small-investor" label could mask non-retail owners that fall below mandatory disclosure thresholds. To address this concern, we design a validation exercise that hinges on an observable, market-wide proxy for retail appetite. A large literature shows that individual investors concentrate in stocks with certain features— high lottery-payoff skewness (Bali et al., 2011), high idiosyncratic volatility (Han and Kumar, 2013), lower firm age (Laarits and Sammon, 2025) and high turnover (Barber and Odean, 2008) are standard flags. We construct these signals based on the prior work and sort the universe into *High-Retail* and *Low-Retail* portfolios each quarter.

We replicate the variance decomposition across six partitions: (i) High-Retail versus Low-Retail stocks, and (ii) within each group, the contributions of Small Investors, Financial Intermediaries (Banks, Pension Funds, and Insurers), and Delegated Asset Management (Mutual Funds and Investment Advisors). This design provides a sharp falsification test. A dominance of Small Investors in High-Retail stocks—but not in Low-Retail stocks—would support our interpretation that the residual primarily reflects household activity. In contrast, similar contributions across investor types would suggest misclassification, implying that the

analysis uses a universe that is about 3% smaller. This difference occurs because not all stocks have the full set of characteristics from Freyberger et al. (2020). As a result, residual ownership by small investors is available for 97% of the stocks.

residual may include unreported institutional investors.

Table C.3 compares the variance share of *Small Investors* (first column) with that of *Other Investors*. The second column reports the difference relative to Financial Intermediaries (Banks, Pension Funds, and Insurers), and the third column reports the difference relative to Delegated Asset Management (Mutual Funds and Investment Advisors), after sorting stocks by four attributes commonly associated with retail trading activity. Across all sorts, the gap (Small−Other)—regardless of which investor group serves as the benchmark—widens sharply in the high-retail quintile, confirming that the residual category behaves as expected for a household proxy.

Panels A and B examine idiosyncratic volatility and maximum daily return (a proxy for lottery-like payoffs). Among low-volatility or low-lottery stocks, the gap between Small and Other Investors is modest—14% and 9%, respectively, when compared to Delegated Asset Management. In the high-volatility or high-lottery quintiles, however, the gap widens to 23% and 10%, driven entirely by a decline in the "Other" share. Small investors are thus more active precisely where risk-seeking motives are strongest.

Panel C sorts stocks by firm age (with higher values indicating younger firms). The share of Small Investors rises from mature to young firms, while the share of Other Investors declines—from positive to near zero—widening the gap in young firms. Panel D repeats the exercise for turnover. Here, the Small-Investor share remains roughly flat across quintiles, but the gap widens substantially (from 16% to 25% relative to Intermediaries, and from –23% to 11% relative to Delegated Asset Management), driven by a decline in participation by institutions. These patterns are consistent with the well-documented retail tilt toward "attention-grabbing" stocks.

In every test the interaction runs in the expected direction: small investors influence intensifies exactly where retail-heavy characteristics concentrate, while institutional influence either fades or turns negative. The systematic nature of these shifts makes it unlikely that our residual is hiding a large pool of unreported institutions; instead, it behaves like a genuine proxy for individual investors.

C.3 Definition of the "Residual" Component

Our portfolio identity in (13) implies a residual that reconciles the realized portfolio log return with the sum of component–induced log returns:

$$u_t \;\equiv\; r_t \;-\; \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}(k)} r_t^{(k_i)}. \tag{C.4}$$

This residual has two sources.

(i) Model reconciliation at the stock level. Because investors' demand functions are estimated without imposing exact market clearing on the last investor, the model–implied equilibrium price can differ slightly from the observed price after all inputs are updated. At the stock level, this gap is the reconciliation term $\boldsymbol{\psi}_t$ in (10), defined by

$$\boldsymbol{\psi}_t \;=\; \mathbf{p}_t \;-\; \mathbf{g}\big(\mathbf{s}_t, \mathbf{X}_t, \mathbf{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_t; \mathcal{J}_t\big). \tag{C.5}$$

Aggregating these stock–level gaps into portfolio space contributes to u_t.

(ii) Log–aggregation (transformation) effect. While log returns add at the stock level by construction of (9), portfolio aggregation involves a nonlinear log of a weighted sum of

simple returns:

$$r_t \;=\; \log\Big(1+\sum_{n=1}^{N} w_{n,t-1}\big(e^{\,r_{n,t}}-1\big)\Big) \;\neq\; \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}(k)} \log\big(1+R_t^{(k_i)}\big), \tag{C.6}$$

where $r_{n,t}$ is the total log return on stock n and $R_t^{(k_i)}$ is defined in (12). The difference between the two sides is a transformation term that is also captured by u_t.

C.4 Bottom-up Approach to Variance Decomposition

We decompose variance by investor and motive using a bottom-up procedure. Fix a motive $k \in \{\mathcal{J}, \mathbf{A}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, \mathbf{s}, \mathbf{X}, v\}$ and an investor type i.

1. Compute $R_t^{(k_i)}$. This is the simple portfolio return induced by updating motive k for investor i:

$$R_t^{(k_i)} = \sum_{n=1}^{N} w_{n,t-1}\big(e^{\,\Delta p_{n,t}(k_i)}-1\big), \tag{C.7}$$

where $w_{n,t-1}$ is the portfolio weight of stock n at $t-1$, and $\Delta p_{n,t}(k_i)$ is the stock-level log price change attributed to the (k,i) component (for dividends, define $\Delta p_{n,t}(v) \equiv v_{n,t}$).

2. Run the regression. For each (k,i), estimate the variance share from

$$\log\big(1+R_t^{(k_i)}\big) = a + b_{k_i}\,\log\big(1+R_t\big) + \varepsilon_t, \tag{C.8}$$

and record b_{k_i}.

3. Summarize across investors. The motive–level effect is the sum of investor–level gains:

$$b_k = \sum_i b_{k_i}. \tag{C.9}$$

Equivalently,

$$b_k^{\text{bottom-up}} = \frac{\mathrm{Cov}\Big(\sum_{n=1}^N w_{n,t-1}\Big[\sum_i e^{\Delta p_{n,t}(k_i)} - I\Big],\ \log\big(1+R_t\big)\Big)}{\mathrm{Var}\big(\log\big(1+R_t\big)\big)}, \tag{C.10}$$

where I is the number of investor types and R_t is the realized simple portfolio return.

Aggregate-first for comparison. Alternatively, aggregate investor effects within motive k first and then estimate the variance decomposition:

$$\log\big(1+R_t^{(k)}\big) = a + b_k \log\big(1+R_t\big) + \varepsilon_t, \qquad R_t^{(k)} = \sum_{n=1}^N w_{n,t-1}\left(\exp\Big\{\sum_i \Delta p_{n,t}(k_i)\Big\} - 1\right), \tag{C.11}$$

so that

$$b_k^{\text{aggregate-first}} = \frac{\mathrm{Cov}\Big(\sum_{n=1}^N w_{n,t-1}\Big[\exp\big\{\sum_i \Delta p_{n,t}(k_i)\big\} - 1\Big],\ \log\big(1+R_t\big)\Big)}{\mathrm{Var}\big(\log\big(1+R_t\big)\big)}. \tag{C.12}$$

With two investors, the difference between the "bottom-up" and "aggregate-firs" estimates comes from $e^{\Delta p_{n,t}(k_1)}+e^{\Delta p_{n,t}(k_2)}-2$ versus $e^{\Delta p_{n,t}(k_1)+\Delta p_{n,t}(k_2)}-1$; when $\Delta p_{n,t}(k_1), \Delta p_{n,t}(k_2) > 0$, the aggregate-first measure is larger because $e^{x+y} > e^x + e^y - 1$ for $x, y > 0$. This inflates the share attributed to motive k and compresses the residual. We therefore adopt the bottom-up approach throughout the paper because it provides a more conservative decomposition.

C.5 Effects Over Varying Market Conditions

To assess whether small investors' influence reflects episodic trading during turbulent or mispriced markets, we re-estimate the variance decompositions after splitting the sample along two market-condition dimensions: (i) high vs. low volatility, and (ii) extreme vs.

less extreme valuation spreads. If small investors trade mainly in stressed conditions, their impact should fall in calm and low-spread regimes; similar magnitudes across all subsamples, however, would indicate a structural motive rather than opportunistic timing.

Panel A of Table C.5 shows that small investors remain the largest contributors in both volatility regimes, though their share declines from 52% in calm periods to 31% during high-volatility quarters. In contrast, mutual funds and advisors become more influential when volatility rises. Panel B indicates that demand for stock characteristics is stronger in turbulent than in calm markets (46% vs. 29%), suggesting that, in volatile markets, investors tend to overreact to public signals embedded in observable firm characteristics. Latent demand and industry-specific demand also increase during volatile periods (38% vs. 32% and 13% vs. 6%, respectively), consistent with greater sentiment-driven and sector-specific trading when market uncertainty is elevated.

Table C.6 partitions the sample into high-valuation-spread quarters—those in the top half of the cross-sectional range of log book-to-market —and low-spread quarters in the bottom half. Small investors remain dominant but are more influential when valuations are dispersed (40% in low-spread vs. 33% in high-value spread quarters), while mutual funds and advisors together are stronger in high-spread periods (39% vs. 18%). Demand for characteristics stays prominent (42% vs. 49%), but latent demand rises sharply in high-spread regimes (42% vs. 14%), consistent with stronger sentiment-driven trading when valuations diverge.

C.6 Variance Decomposition into the Effects of Cash Flow and Discount Rate News

Below, we outline the method used to calculate cash flow and discount rate news for anomaly portfolios. Our approach closely follows the method proposed by Lochstoer and Tetlock

(2020). We assume that the expected log returns for stock i at time t are linear functions of market-adjusted characteristics, $X_{i,t}^{ma}$, and a vector of aggregate characteristics, X_t^{agg}:

$$E_t[r_{i,t+1}] = \delta_0 + \delta_1^T X_{i,t}^{ma} + \delta_2^T X_t^{agg}. \tag{C.13}$$

We utilize the same set of firm characteristics as in Lochstoer and Tetlock (2020): log of book equity to market equity, five-year change in log market equity, log profitability, five-year log asset growth, and six-month momentum.

To obtain the return decomposition, we first estimate an aggregate VAR on the value-weighted market log return r_t:

$$Z_{t+1} = \mu^{agg} + A^{agg} Z_t + \varepsilon_{t+1}^{agg}, \tag{C.14}$$

where Z_t is a vector that includes the value-weighted market log return r_t and the value-weighted firm characteristics X_t^{agg}.

The aggregate discount rate and cash flow shocks, DR_{t+1}^{agg} and CF_{t+1}^{agg}, are computed using the standard VAR formula:

$$\begin{aligned} DR_{t+1}^{agg} &= e_1' \kappa A^{agg} \left(I_{K^{\text{agg}}} - \kappa A^{agg}\right)^{-1} \varepsilon_{t+1}^{agg}, \\ CF_{t+1}^{agg} &= e_1' \left(I_{K^{agg}} + \kappa A^{agg} \left(I_{K^{agg}} - \kappa A^{agg}\right)^{-1}\right) \varepsilon_{t+1}^{agg}, \end{aligned} \tag{C.15}$$

where e_1 is a $K^{\text{agg}} \times 1$ column vector with one as its first element and zeros elsewhere, $I_{K^{\text{agg}}}$ is a $K^{\text{agg}} \times K^{\text{agg}}$ identity matrix, and $\kappa = 0.95$.

To capture cross-sectional persistence in characteristics, we deviate from Lochstoer and Tetlock (2020) and follow Chen et al. (2013) by assuming that the slope δ_1 in Equation (C.13) varies by firm size, and that there is a firm-fixed effect in the data-generating process captured by $\delta_{0,i}$. Thus, we group stocks into ten deciles based on their unconditional mean of the cross-sectional size percentile using breakpoints based on all stocks, rather than the NYSE breakpoints, and estimate the market-adjusted panel VAR with stock fixed effects for all stocks i in a given size decile:

$$Z_{i,t+1} = \mu_i^{ma} + A^{ma} Z_{i,t} + \varepsilon_{i,t+1}, \tag{C.16}$$

where $Z_{it} = [r_{it}^{ma}; X_{it}^{ma}]$ is a $K^{ma} \times 1$ vector, r_{it}^{ma} is the log return for stock i adjusted by the value-weighted log market return, and $X_{i,t}^{ma}$ are market-adjusted firm characteristics. Then, for each decile group, we calculate the discount rate and cash flow shocks at the stock level, $DR_{i,t+1}^{ma}$ and $CF_{i,t+1}^{ma}$, as:

$$\begin{aligned} DR_{t+1}^{ma} &= e_1' \kappa A^{ma} \left(I_{K^{\mathrm{ma}}} - \kappa A^{ma}\right)^{-1} \varepsilon_{i,t+1}, \\ CF_{t+1}^{ma} &= e_1' \left(I_{K^{ma}} + \kappa A^{ma} \left(I_{K^{ma}} - \kappa A^{ma}\right)^{-1}\right) \varepsilon_{i,t+1}. \end{aligned} \tag{C.17}$$

We then sum the aggregate news and market-adjusted news to obtain the stock-level news:

$$\begin{aligned} DR_{i,t} &= DR_t^{agg} + DR_{i,t}^{ma}, \\ CF_{i,t} &= CF_t^{agg} + CF_{i,t}^{ma}. \end{aligned} \tag{C.18}$$

To maintain consistency with our main analysis, we transform the CF and DR shocks from log returns to shocks in simple returns. We follow the approach of Lochstoer and Tetlock (2020), who derive an expression for shocks to simple returns using the Taylor approximation:

$$\begin{aligned} CF_{i,t+1}^{\text{simple}} &\equiv \exp\left(E_t r_{i,t+1}\right) \left\{ CF_{i,t+1} + \frac{1}{2} \left(CF_{i,t+1}\right)^2 \right\}, \\ DR_{i,t+1}^{\text{simple}} &\equiv \exp\left(E_t r_{i,t+1}\right) \left\{ DR_{i,t+1} - \frac{1}{2} \left(DR_{i,t+1}\right)^2 \right\}. \end{aligned} \tag{C.19}$$

The portfolio-level innovation in return for a portfolio p with weights $w_{i,t}$ in stock i at time t is calculated as:

$$\tilde{R}_{p,t+1} \equiv R_{p,t+1} - \sum_i w_{i,t}^p \exp\left(E_t r_{i,t+1}\right) \approx CF_{p,t+1}^{simple} - DR_{p,t+1}^{simple} + CFDR_{p,t+1}^{cross}, \tag{C.20}$$

where

$$\begin{aligned}
CF_{p,t+1}^{simple} &= \sum_i w_{i,t}^p CF_{i,t+1}^{\text{simple}}, \\
DR_{p,t+1}^{simple} &= \sum_i w_{i,t}^p DR_{i,t+1}^{\text{simple}}, \\
CFDR_{p,t+1}^{cross} &= \sum_i w_{i,t}^p \left(\exp\left(E_t r_{i,t+1}\right) CF_{i,t+1} DR_{i,t+1}\right).
\end{aligned} \tag{C.21}$$

In calculating variance, we follow the approach of Lochstoer and Tetlock (2020), which omits the final term in Equation (C.20). They demonstrate that for simple returns, the covariance of return innovations with CF and DR shocks does not fully explain the overall variance due to this omitted term. However, Lochstoer and Tetlock (2020) show that the impact of this term is minimal.

To incorporate our investor-motive breakdown into the CF-DR decomposition, we follow the previously described process, replacing the log return for stock i with the component resulting from a change in motive k for investor i. This gives the investor-by-motive innovation to the simple portfolio return, $\tilde{R}_{p,t}^{(k_i)}$. Following our main approach in Section 2.3.2, we apply a log-transformation to all return innovations, as well as CF and DR shocks. Given that for log-returns we have $\tilde{r}_{p,t} \approx \sum_k \sum_i \tilde{r}_{p,t}^{(k_i)} \approx cf_{p,t} - dr_{p,t}$, we can represent the variance in log return innovations as the sum of the covariances of log component innovations with CF and DR shocks:

$$Var(\tilde{r}_{p,t}) = \sum_k \sum_i \text{Cov}\left(\tilde{r}_{p,t}^{(k_i)}, cf_{p,t}\right) + \sum_k \sum_i \text{Cov}\left(\tilde{r}_{p,t}^{(k_i)}, -dr_{p,t}\right) + \text{Cov}\left(u_t, \tilde{r}_{p,t}\right). \quad \text{(C.22)}$$

In this equation, u_t represents the model's total residual, which includes both the effects of the omitted final term in Equation (C.20) and the impact of the log-transformation. By normalizing both sides of Equation (C.22) with $\text{Var}(\tilde{r}_{p,t})$, we derive a three-part variance decomposition. This effectively divides the total variance into contributions from motives, investors, and CF-DR shocks.

C.7 Relation Between Latent Demand and Fundamentals

C.7.1 Does Latent Demand Contain Information About Future Fundamentals?

We use machine learning to assess latent demand's ability to forecast future fundamentals. First, we train a benchmark LightGBM model at the stock level to predict the next quarter's characteristics using only contemporaneous characteristics. We then augment the baseline model by including latent demand from specific investor types as an additional predictor and compare the performance of the augmented model with the benchmark model. If the model incorporating latent demand demonstrates superior predictive power, it suggests that latent demand contains information about future fundamentals.

We use the Gradient Boosting Regression Tree Model (LightGBM) to forecast firm characteristics $Y_{t+1}(n)$ one year ahead, focusing on the inputs to our demand system: book equity, market equity, profitability, investment, accruals, dividend-to-book ratio, and market beta. Rather than predicting book and market equity separately, however, we forecast the book-to-market ratio. As Vuolteenaho (2000, 2002) demonstrates, the book-to-market ratio is linked to future cash-flow fundamentals, specifically earnings over book equity, making it a more effective proxy for fundamental value than each component on its own.

In our benchmark model, we predict stock characteristics using only the contemporaneous characteristics $X_t(n)$ as features. We then enhance this model by including latent demand as an additional feature. Since we train the model at the stock level, we first aggregate the latent demand from all investors in group g at the stock level:

$$\epsilon_t^g(n) = \sum_{i \in g} w_{i,t}(n) \epsilon_{i,t}(n). \tag{C.23}$$

We then incorporate $\epsilon_t^g(n)$ into the predictive model of stock characteristics:

$$Y_{t+1}(n) = g_t(X_t(n); \epsilon_t^g(n)). \tag{C.24}$$

We train the model in Equation (C.24) at the stock level using the rolling method from Gu et al. (2020). The 1999-2022 sample data is divided into a 7-year training period (1999-2005), a 3-year validation period (2006-2008), and a 14-year out-of-sample testing period (2009-2022). We refit the model annually, expanding the training sample by one year each time and using the next three years as the validation set. Since most dependent variables are updated annually, we only use the June sample for annual predictions. All variables are rank-normalized cross-sectionally following the method in Gu et al. (2020). We use grid search to determine the optimal learning rate and maximum depth from the validation set.

We compare the models' performances using two metrics. Mean squared error (MSE) measures predictive accuracy; a negative difference in MSE indicates that the model with latent demand is more accurate in forecasting future characteristics. We use Diebold-Mariano (DM) t-statistics to test the statistical significance of MSE differences. The DM statistic is determined by regressing the difference in squared prediction errors on a constant, using double-clustered standard errors. Out-of-sample R^2 (R^2_{OOS}) measures explanatory power; a positive difference indicates greater explanatory power for the model with latent demand. The R^2_{OOS} is calculated as one minus the ratio of the squared prediction error to the squared realized value. We consider the model with latent demand superior if it shows significantly higher predictive accuracy and explanatory power.

Panel (a) of Table C.7 shows that adding latent demand to the forecasting models rarely improves out-of-sample performance. For market beta, dividend-to-book equity, asset growth, and accruals, the differences in MSE are positive and the R^2_{OOS} differences are negative across all investor types, indicating that the augmented models perform worse than the benchmark. The deterioration is especially pronounced for market beta, where R^2_{OOS} falls by about 1.5–2%, and for accruals, where R^2_{OOS} decreases by roughly 18%.

The only characteristics that benefit from including latent demand are book-to-market and operating profits-to-book equity. For both, the MSE differences are negative and statistically significant, with small R^2_{OOS} gains for book-to-market (up to about 0.2%) and more meaningful improvements for operating profitability (around 2.5–2.7%). Overall, latent demand provides limited incremental predictive power for future fundamentals and, in several cases, substantially reduces forecasting accuracy.

C.7.2 Does Latent Demand Reflect Historical Fundamental Information?

We next establish a benchmark machine learning model to predict future latent demand for each investor type using current latent demand. We then test whether adding past stock characteristics enhances the model's predictive accuracy. For these tests, we draw on a broad set of 209 stock characteristics from Open Source Asset Pricing (Chen and Zimmermann, 2022), which includes our baseline characteristics.[30] By using this extensive set of fundamentals, we ensure the model has access to as much stock-related information as possible, giving the fundamentals-augmented model a substantial opportunity to improve its prediction of latent demand.

When predicting latent demand, we use an approach similar to the one adopted in the previous section. In particular, we train the Gradient Boosting Regression Tree Model (LightGBM) to predict 1-year ahead investor group-stock-level latent demand, $\epsilon^g_{t+1}(n)$, defined as

[30] See https://www.openassetpricing.com.

in Equation (C.23). In the benchmark model, we use only contemporaneous latent demand as a feature. In the augmented model, we include stock characteristics $X_t(n)$ as additional features. The augmented model is then expressed as:

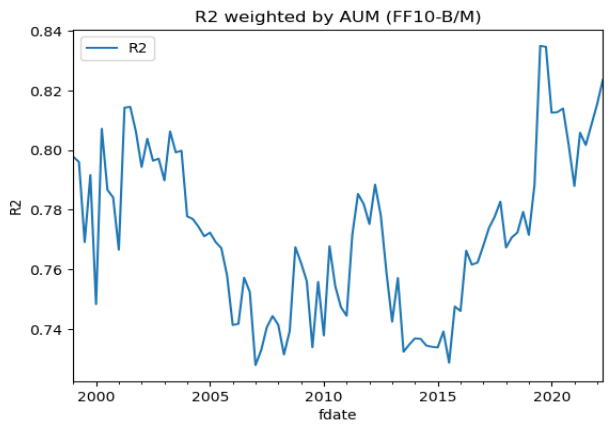
$$\epsilon^g_{t+1}(n) = g_t(X_t(n); \epsilon^g_t(n)). \tag{C.25}$$

where we include all firm characteristics from the Open Source Asset Pricing Dataset in $X_t(n)$. All other steps follow the same procedure as in the previous subsection where we predict stock characteristics. In particular we compute the difference between the augmented and benchmark models for both R^2_{OOS} and MSE. A negative R^2_{OOS} difference and a positive MSE difference indicate that the benchmark model is superior, implying the augmented model with additional characteristics lacks explanatory power.

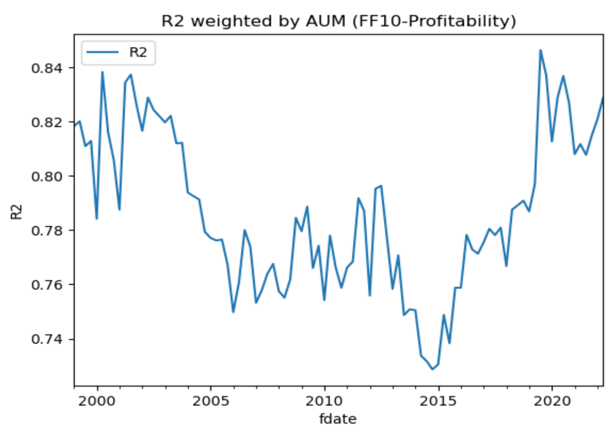
Panel (b) of Table C.7 indicates that adding past characteristics does not enhance forecast accuracy. Across all investor groups, we observe consistent positive and statically significant differences in MSE and negative differences in R^2_{OOS}. These results suggest that latent demand is unrelated to historical fundamentals, strengthening its interpretation as non-fundamentals-based trading.

Figure C.1: Time-Series of the Average AUM-Weighted R^2

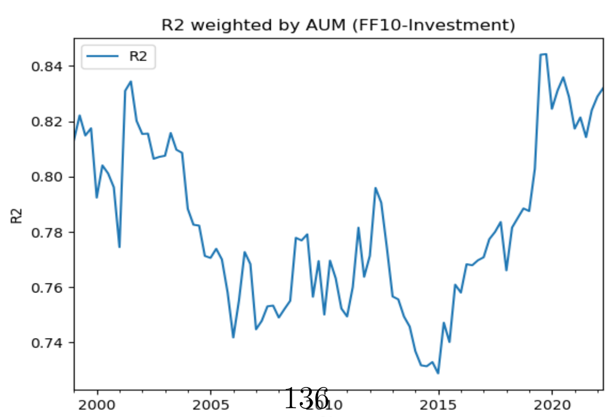
Panel (a) shows the industry-value series, Panel (b) shows the industry-profitability series, and Panel (c) shows the industry-investment series. Each subfigure plots the AUM-weighted average R^2 over time for the specified sorting.



(a) Industry-Value



(b) Industry-Profitability



(c) Industry-Investment

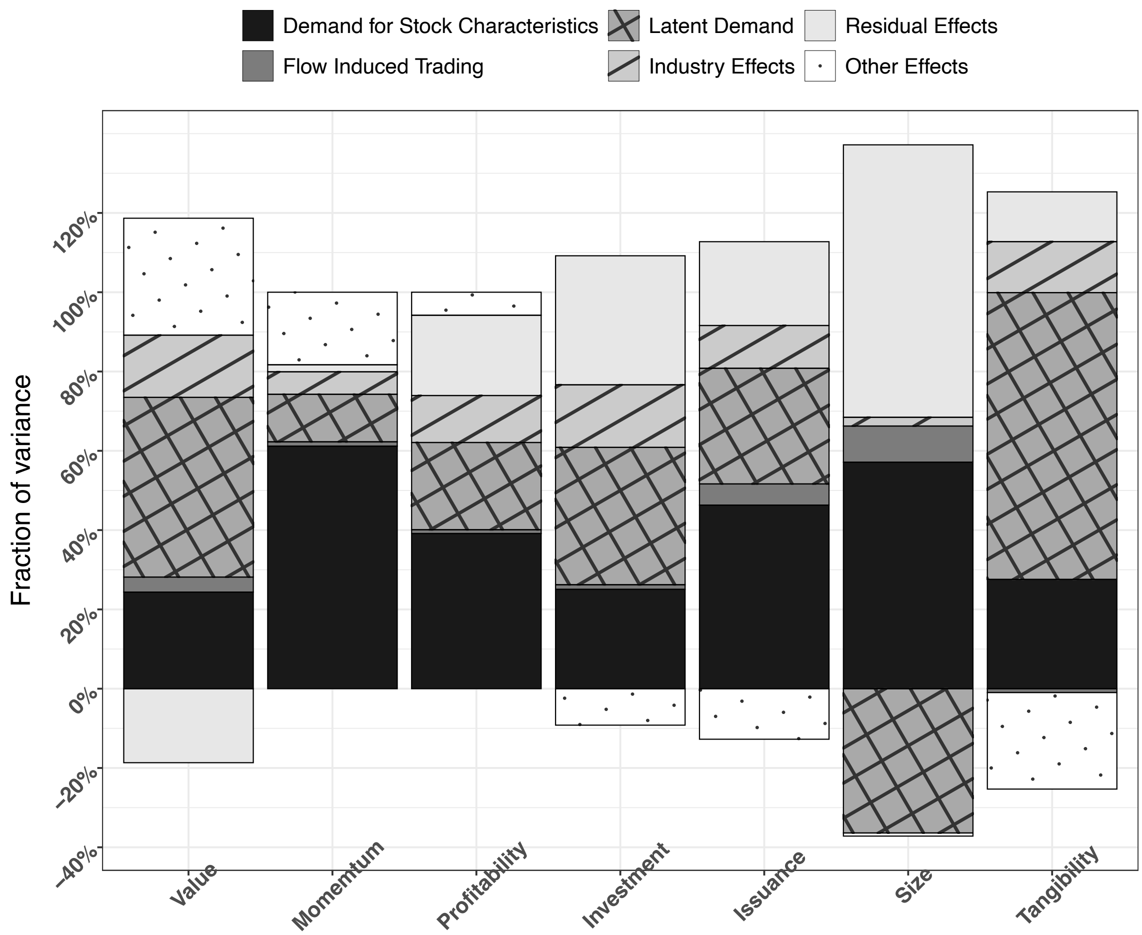


Figure C.2: Variance Decomposition by Trading Motives: Long Leg
In this figure, we report the decomposition of return variation in the long leg of the anomaly. Each bar represents a specific anomaly group. Appendix Table A.1 lists the anomalies included in each group. We aggregate individual anomalies into groups using the method outlined in Section 2.3.2. The bars illustrate the proportions of total anomaly return variance attributed to demand for stock characteristics, latent demand, flow-induced trading, other effects, and residual effects. "Other Effects" encompass changes in shares outstanding, changes in AUM driven by returns, changes in stock characteristics, and dividend yield. "Residual Effects" include portfolio adjustment impacts and model error. Detailed information can be found in Section 2. The total sum of all effects is 100%.

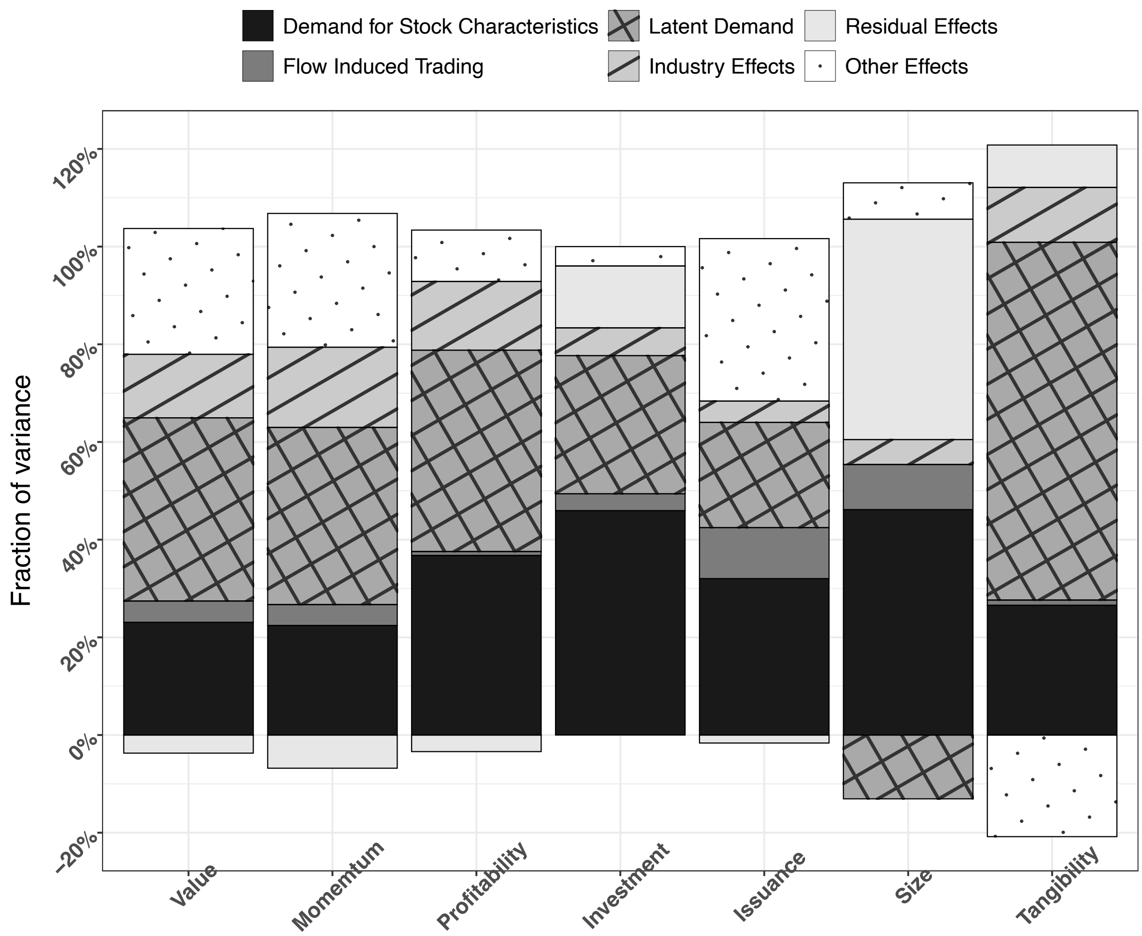


Figure C.3: Variance Decomposition by Trading Motives: Short Leg
In this figure, we report the decomposition of return variation in the short leg of the anomaly. Each bar represents a specific anomaly group. Appendix Table A.1 lists the anomalies included in each group. We aggregate individual anomalies into groups using the method outlined in Section 2.3.2. The bars illustrate the proportions of total anomaly return variance attributed to demand for stock characteristics, latent demand, flow-induced trading, other effects, and residual effects. "Other Effects" encompass changes in shares outstanding, changes in AUM driven by returns, changes in stock characteristics, and dividend yield. "Residual Effects" include portfolio adjustment impacts and model error. Detailed information can be found in Section 2. The total sum of all effects is 100%.

Table C.1: Fraction of Anomaly Portfolio Dollar Value Held by Investor Group

This table reports the average percentage of the dollar value of anomaly portfolios held by various investor groups. A value of 100% means that the investor group holds the entire dollar value of the anomaly portfolio. Detailed computations are provided in Appendix Section C.1. Panel (a) presents the results for the long leg, while Panel (b) shows the results for the short leg.

Period	Anomaly	Small Investors	Mutual Funds	Investment Advisors	Banks	Insurance Companies	Pension Funds	Other Investors	Short Sellers	Insiders	Blockholders
1999-2005	Value	38%	27%	11%	14%	4%	3%	1%	2%	0%	0%
	Momentum	35%	29%	11%	15%	4%	3%	1%	2%	0%	0%
	Profitability	38%	27%	11%	14%	4%	3%	1%	2%	0%	0%
	Investment	37%	27%	11%	14%	4%	3%	1%	2%	0%	0%
	Issuance	39%	26%	10%	16%	4%	3%	1%	2%	0%	0%
	Size	47%	25%	12%	8%	2%	2%	1%	2%	0%	0%
	Tangibility	38%	28%	10%	14%	4%	3%	1%	2%	0%	0%
2006-2010	Value	27%	29%	18%	14%	4%	3%	2%	3%	1%	0%
	Momentum	24%	30%	19%	14%	3%	3%	2%	4%	1%	0%
	Profitability	28%	28%	18%	14%	3%	3%	2%	3%	1%	0%
	Investment	27%	29%	18%	14%	3%	3%	2%	3%	1%	0%
	Issuance	31%	27%	17%	15%	3%	3%	2%	2%	0%	0%
	Size	27%	28%	20%	10%	2%	2%	2%	7%	2%	0%
	Tangibility	24%	30%	19%	14%	3%	3%	2%	4%	1%	0%
2011-2015	Value	29%	27%	20%	14%	2%	3%	2%	3%	1%	0%
	Momentum	25%	28%	21%	15%	2%	2%	2%	3%	1%	0%
	Profitability	27%	27%	21%	14%	2%	2%	2%	3%	1%	0%
	Investment	27%	27%	20%	14%	2%	3%	2%	3%	1%	0%
	Issuance	29%	27%	20%	15%	2%	3%	2%	2%	0%	0%
	Size	30%	26%	21%	9%	1%	1%	2%	6%	2%	0%
	Tangibility	25%	28%	22%	14%	2%	2%	2%	4%	1%	0%
2016-2022	Value	26%	28%	21%	14%	2%	2%	3%	2%	1%	0%
	Momentum	26%	28%	21%	14%	2%	2%	3%	3%	1%	0%
	Profitability	26%	28%	21%	14%	2%	2%	3%	3%	1%	0%
	Investment	27%	28%	21%	14%	2%	2%	3%	3%	1%	0%
	Issuance	30%	27%	20%	14%	2%	2%	3%	2%	1%	0%
	Size	23%	28%	24%	9%	1%	1%	3%	6%	3%	1%
	Tangibility	28%	27%	21%	13%	2%	2%	3%	3%	1%	0%

(a) Long Leg

Period	Anomaly	Small Investors	Mutual Funds	Investment Advisors	Banks	Insurance Companies	Pension Funds	Other Investors	Short Sellers	Insiders	Blockholders
1999-2005	Value	37%	28%	10%	15%	4%	3%	1%	2%	0%	0%
	Momentum	39%	27%	11%	13%	4%	3%	1%	2%	0%	0%
	Profitability	39%	26%	10%	15%	4%	3%	1%	2%	0%	0%
	Investment	37%	27%	10%	15%	4%	3%	1%	2%	0%	0%
	Issuance	36%	28%	11%	14%	5%	3%	1%	2%	0%	0%
	Size	40%	25%	9%	16%	4%	3%	1%	1%	0%	0%
	Tangibility	37%	27%	10%	15%	4%	3%	1%	2%	0%	0%
2006-2010	Value	26%	29%	19%	14%	3%	3%	2%	4%	1%	0%
	Momentum	26%	29%	20%	13%	3%	3%	2%	5%	1%	0%
	Profitability	29%	28%	18%	14%	3%	3%	2%	3%	1%	0%
	Investment	26%	29%	19%	14%	3%	3%	2%	3%	1%	0%
	Issuance	24%	30%	20%	14%	3%	3%	2%	4%	1%	0%
	Size	32%	27%	16%	15%	3%	3%	1%	2%	0%	0%
	Tangibility	25%	30%	19%	14%	3%	3%	2%	3%	1%	0%
2011-2015	Value	26%	28%	21%	14%	2%	2%	2%	3%	1%	0%
	Momentum	27%	27%	21%	13%	2%	2%	2%	4%	1%	0%
	Profitability	28%	27%	20%	14%	2%	2%	2%	3%	1%	0%
	Investment	27%	28%	21%	14%	2%	2%	2%	3%	1%	0%
	Issuance	26%	28%	21%	13%	2%	2%	2%	4%	1%	0%
	Size	30%	26%	19%	15%	2%	3%	2%	2%	0%	0%
	Tangibility	24%	29%	22%	15%	2%	3%	2%	3%	1%	0%
2016-2022	Value	26%	28%	22%	14%	2%	2%	3%	3%	1%	0%
	Momentum	23%	28%	23%	13%	2%	2%	3%	4%	1%	0%
	Profitability	26%	28%	21%	14%	2%	2%	3%	3%	1%	0%
	Investment	26%	28%	21%	14%	2%	2%	3%	3%	1%	0%
	Issuance	21%	30%	24%	13%	2%	2%	3%	4%	2%	0%
	Size	30%	27%	20%	14%	2%	2%	3%	2%	1%	0%
	Tangibility	24%	29%	23%	14%	2%	2%	3%	2%	1%	0%

(b) Short Leg

Table C.2: Percentage of Stocks Within Anomaly Portfolio Held by Investor Groups

This table shows the average percentage of stocks within each anomaly group held by various investor groups. A value of 100% indicates that the investor group holds every stock assigned to the anomaly portfolio. The detailed computations are provided in Appendix Section C.1. Panel (a) presents the results for the long leg, while Panel (b) presents the results for the short leg.

	Small Investors	Mutual Funds	Investment Advisors	Banks	Insurance Companies	Pension Funds	Other Investors	Short Sellers	Insiders	Blockholders
Value	96%	89%	86%	89%	66%	65%	51%	63%	55%	71%
Momentum	100%	93%	92%	94%	81%	75%	64%	65%	58%	73%
Profitability	96%	88%	86%	88%	67%	66%	52%	61%	57%	72%
Investment	96%	85%	82%	86%	59%	59%	45%	59%	57%	73%
Issuance	96%	91%	89%	91%	75%	73%	59%	67%	58%	68%
Size	97%	84%	82%	85%	55%	55%	40%	55%	56%	73%
Tangibility	95%	86%	84%	87%	62%	62%	47%	59%	57%	75%

(a) Long Leg

	Small Investors	Mutual Funds	Investment Advisors	Banks	Insurance Companies	Pension Funds	Other Investors	Short Sellers	Insiders	Blockholders
Value	95%	88%	87%	89%	69%	67%	53%	59%	58%	72%
Momentum	100%	89%	89%	91%	73%	68%	57%	61%	57%	73%
Profitability	96%	88%	86%	89%	68%	66%	52%	61%	58%	68%
Investment	96%	91%	90%	91%	76%	74%	61%	67%	57%	70%
Issuance	96%	88%	87%	89%	68%	66%	51%	61%	59%	74%
Size	99%	100%	100%	99%	99%	98%	95%	89%	56%	57%
Tangibility	96%	89%	87%	89%	69%	67%	52%	61%	57%	70%

(b) Short Leg

Table C.3: Differences in Variance Shares Across Stocks and Investors

Each panel reports, for the lowest (Low) and highest (High) quintiles of the sorting variable, the share of variance explained by Small Investors. The second column shows the difference in this share relative to Financial Intermediaries (i.e., Banks, Pension Funds, and Insurers), and the third column shows the difference relative to Delegated Asset Management (i.e., Mutual Funds and Investment Advisors).

(a) Panel A: Idiosyncratic Volatility

Quantile	Small Investors (SI)	Diff: SI - Fin. Intermediaries	Diff: SI - Del. Asset Mngt.
Low Retail	47%	33%	14%
High Retail	44%	39%	23%

(b) Panel B: Maximum Daily Return

Quantile	Small Investors (SI)	Diff: SI - Fin. Intermediaries	Diff: SI - Del. Asset Mngt.
Low Retail	50%	16%	9%
High Retail	38%	29%	10%

(c) Panel C: Firm Age (High = Younger)

Quantile	Small Investors (SI)	Diff: SI - Fin. Intermediaries	Diff: SI - Del. Asset Mngt.
Low Retail	-0.1%	-47%	-40%
High Retail	60%	71%	58%

(d) Panel D: Turnover

Quantile	Small Investors (SI)	Diff: SI - Fin. Intermediaries	Diff: SI - Del. Asset Mngt.
Low Retail	34%	16%	-23%
High Retail	35%	25%	11%

Table C.4: AUM-Weighted Time-Series Average R-Squared Across Investors

This table reports the AUM-weighted mean of the time-series average R^2 for each investor type, separately for Industry-Value (Panel A), Industry-Profitability (Panel B), and Industry-Investment (Panel C) portfolios.

	Small Investors	Banks	Insurance Companies	Investment Advisors	Mutual Funds	Pension Funds	Other	Short Sellers	Insiders	Blockholders
Panel A: Industry-Value										
AUM-weighted Mean	95%	84%	73%	47%	71%	88%	55%	66%	54%	71%
Panel B: Industry-Profitability										
AUM-weighted Mean	96%	86%	75%	49%	73%	87%	56%	67%	57%	75%
Panel C: Industry-Investment										
AUM-weighted Mean	95%	86%	74%	49%	72%	88%	56%	69%	56%	78%

Table C.5: Variance Decomposition Across High- and Low-Volatility Periods

Panel A reports the variance share attributable to each investor group during high- and low-volatility regimes. Panel B reports the variance share attributable to each trading motive during those regimes. Rows correspond to anomaly portfolios.

Panel A: Investor-Level Shares

Anomaly	Small Investors		Investment Advisors		Mutual Funds		Other Investors	
	High Vol	Low Vol	High Vol	Low Vol	High Vol	Low Vol	High Vol	Low Vol
Value	20%	39%	28%	-1%	50%	24%	38%	-14%
Momentum	57%	74%	14%	-29%	21%	10%	10%	-21%
Profitability	24%	48%	16%	-46%	32%	2%	22%	-31%
Investment	30%	-11%	10%	-30%	17%	8%	14%	-25%
Issuance	43%	121%	-1%	13%	40%	25%	66%	33%
Size	12%	45%	12%	-10%	33%	27%	5%	7%
Tangibility	30%	48%	18%	-1%	17%	36%	-3%	-31%
Average	31%	52%	14%	-15%	30%	19%	22%	-12%

Panel B: Trading-Motive Shares

Anomaly	Demand for Characteristics		Flow-Induced Trading		Latent Demand		Industry-specific Demand	
	High Vol	Low Vol	High Vol	Low Vol	High Vol	Low Vol	High Vol	Low Vol
Value	33%	1%	-6%	-1%	84%	53%	25%	7%
Momentum	62%	53%	-4%	2%	33%	-21%	11%	0%
Profitability	49%	57%	-5%	1%	38%	51%	16%	6%
Investment	32%	-16%	2%	-7%	27%	17%	10%	9%
Issuance	61%	28%	17%	22%	67%	17%	2%	3%
Size	39%	61%	2%	6%	11%	87%	10%	7%
Tangibility	42%	18%	-2%	-15%	5%	22%	17%	12%
Average	46%	29%	1%	1%	38%	32%	13%	6%

Table C.6: Variance Decomposition Across High- and Low-Value-Spread Periods

Panel A reports variance shares attributable to each investor group during high- and low-value-spread regimes. Panel B reports variance shares attributable to each trading motive during those regimes. Rows correspond to anomaly portfolios.

Panel A: Investor-Level Shares

Anomaly	Small Investors		Investment Advisors		Mutual Funds		Other Investors	
	High Spread	Low Spread	High Spread	Low Spread	High Spread	Low Spread	High Spread	Low Spread
Value	24%	39%	25%	6%	46%	8%	37%	11%
Momentum	60%	9%	7%	12%	20%	27%	7%	10%
Profitability	29%	49%	19%	8%	29%	6%	26%	6%
Investment	23%	25%	4%	-5%	20%	11%	17%	7%
Issuance	62%	-8%	-9%	4%	41%	11%	57%	19%
Size	7%	79%	13%	22%	32%	-2%	29%	3%
Tangibility	23%	87%	18%	8%	11%	8%	-5%	13%
Average	33%	40%	11%	8%	28%	10%	24%	10%

Panel B: Trading-Motive Shares

Anomaly	Demand for Characteristics		Flow-Induced Trading		Latent Demand		Industry-specific Demand	
	High Spread	Low Spread	High Spread	Low Spread	High Spread	Low Spread	High Spread	Low Spread
Value	28%	42%	-5%	1%	86%	17%	23%	4%
Momentum	63%	17%	-3%	-2%	25%	47%	10%	-4%
Profitability	52%	52%	-4%	-2%	45%	16%	15%	9%
Investment	23%	24%	1%	-5%	30%	18%	11%	0%
Issuance	60%	33%	19%	11%	70%	-15%	2%	-2%
Size	36%	93%	2%	11%	34%	-15%	9%	12%
Tangibility	32%	85%	-4%	-3%	1%	29%	17%	5%
Average	42%	49%	1%	1%	42%	14%	12%	4%

Table C.7: Predictive Relationships Between Latent Demand and Stock Characteristics

The table reports several key metrics from forecasting 1-year ahead stock characteristics and latent demand using LightGBM machine learning models. We present the differences in mean squared error (MSE), Diebold-Mariano (DM) t-statistics for these differences, and the differences in out-of-sample R^2 (R^2_{OOS}). In Panel (a), we predict stock characteristics and calculate the metrics as differences between the augmented model, which adds contemporaneous latent demand as an additional predictor, and the benchmark model, which only includes the contemporaneous characteristics. In Panel (b), we predict latent demand and calculate the metrics as differences between the augmented model, which adds contemporaneous stock characteristics and the benchmark model, which only includes the contemporaneous latent demand. A negative difference in MSE and a positive difference in R^2_{OOS} indicate the outperformance of the augmented model. Panel (b) uses a comprehensive set of 209 firm characteristics from Open Asset Pricing, including the characteristics used in estimating the demand system.

Panel (a): Predicting Stock Characteristics With and Without Latend Demand

	Small Investors	Mutual Funds	Investment Advisors	Banks	Insurance Companies	Pension Funds	Other	Short-Sellers	Blockholders
Market Beta									
Difference in MSE	56.44	55.97	56.41	55.97	54.23	56.03	55.5	56.46	57.65
DM t-statistic	(3.89)	(3.87)	(3.89)	(3.95)	(4.04)	(3.84)	(4.05)	(3.89)	(3.79)
Diff in R^2_{OOS}	[-1.84%]	[-1.82%]	[-1.83%]	[-1.78%]	[-1.35%]	[-1.61%]	[-1.45%]	[-1.82%]	[-1.79%]
Book-to-Market									
Difference in MSE	-1.69	-1.45	-1.39	-1.36	-1.29	-1.37	-1.41	-1.32	-1.51
DM t-statistic	(-5.18)	(-4.99)	(-4.63)	(-4.53)	(-3.88)	(-4.98)	(-4.41)	(-4.57)	(-4.92)
Diff in R^2_{OOS}	[0.03%]	[0.09%]	[0.03%]	[0.13%]	[0.01%]	[0.19%]	[0.02%]	[0.01%]	[0.02%]
Dividend/Book Equity									
Difference in MSE	0.16	0.17	0.15	0.11	0.04	0.15	0.13	0.15	0.25
DM t-statistic	(0.99)	(1.06)	(0.98)	(0.79)	(0.34)	(0.92)	(0.78)	(0.97)	(1.40)
Diff in R^2_{OOS}	[-0.11%]	[-0.1%]	[-0.09%]	[-0.01%]	[-0.39%]	[-0.07%]	[-0.25%]	[-0.11%]	[-0.35%]
Operating Profits/Book Equity									
Difference in MSE	-0.47	-0.45	-0.46	-0.45	-0.44	-0.47	-0.49	-0.49	-0.52
DM t-statistic	(-3.60)	(-3.42)	(-3.54)	(-3.39)	(-3.90)	(-3.83)	(-3.58)	(-3.70)	(-3.07)
Diff in R^2_{OOS}	[2.7%]	[2.65%]	[2.67%]	[2.6%]	[2.51%]	[2.53%]	[2.59%]	[2.69%]	[2.62%]
Asset Growth									
Difference in MSE	0.85	0.84	0.86	0.89	0.98	0.9	0.73	0.86	1.06
DM t-statistic	(1.88)	(1.81)	(1.88)	(1.94)	(2.08)	(1.89)	(1.49)	(1.89)	(2.08)
Diff in R^2_{OOS}	[-0.27%]	[-0.18%]	[-0.17%]	[-0.09%]	[-0.6%]	[-0.18%]	[-0.41%]	[-0.29%]	[-0.22%]
Accruals									
Difference in MSE	0.39	0.6	0.47	0.48	0.35	0.4	0.46	0.7	0.45
DM t-statistic	(0.69)	(1.09)	(0.85)	(0.87)	(0.67)	(0.78)	(0.81)	(1.35)	(0.81)
Diff in R^2_{OOS}	[-18.62%]	[-18.52%]	[-18.53%]	[-18.37%]	[-17.22%]	[-18.01%]	[-17.8%]	[-18.6%]	[-19.17%]

Table C.7: Predictive Relationships Between Latent Demand and Stock Characteristics (continued)

Panel (b): Predicting Latent Demand With and Without Stock Characteristics

	Small Investors	Mutual Funds	Investment Advisors	Banks	Insurance Companies	Pension Funds	Other	Short-Sellers	Insiders	Blockholders
Difference in MSE	75.84	1.40	0.03	0.28	2.00	0.00	0.02	0.52	0.02	0.00
DM t-statistic	(5.47)	(3.53)	(4.81)	(4.15)	(4.81)	(6.62)	(6.42)	(6.52)	(6.24)	(5.35)
Diff in R^2_{OOS}	[-0.11%]	[-0.12%]	[-0.18%]	[-0.13%]	[-0.15%]	[-0.10%]	[-0.13%]	[-0.72%]	[-0.28%]	[-0.36%]