Estimating Asset Demand: Measurement Error and Predictability

Ryan Cheng*

November 20, 2026

Click here for the latest draft.

Abstract

Previous applications of structural demand systems to asset pricing have found an extremely inelastic stock market. Alongside this result, the use of sign restrictions to ensure downward-sloping demand and the instability of the estimates have raised doubts about the approach's suitability for financial markets. This paper examines whether these features are manifestations of measurement error. Using an estimator consistent with the error structure in portfolio holdings, I obtain naturally downward-sloping demand estimates without imposing sign restrictions. The estimates are stable over time and predict funds' subsequent trading. Estimated elasticities are substantially higher, but the central conclusion survives: the stock market is inelastic.

JEL Codes: G11, G12, G23, C13, C18.
Keywords: Demand-system asset pricing, measurement error, portfolio holdings, demand elasticity, granularity.

*INSEAD. Email: ryan.cheng@insead.edu. I thank Frederico Belo, Manav Chaudhary, Sergei Glebkin, Péter Kondor, Xinyu Liu, Anna Pavlova, Boris Vallée, Malcolm Wardlaw and Motohiro Yogo for valuable feedback. I further thank participants at the INSEAD–Wharton Doctoral Consortium, the Macro Finance Society Workshop and the Northern Finance Association for helpful comments. A previous version of this paper was circulated under the title "Portfolio Granularity and Demand Estimation." All errors are my own.

Introduction

Demand systems estimated from market shares have a long history in applied economics. From discrete-choice models (McFadden, 1974) to differentiated-product demand (Berry et al., 1995; Nevo, 2001), their attraction is not that they provide a complete description of individual behavior. It is that they impose enough structure on high-dimensional choice data to deliver tractable, interpretable objects that can be estimated and taken to counterfactuals. Over time, empirical work has adapted that basic structure to the particular features of different markets.

Recent interest in adapting this approach to financial markets starts with Koijen and Yogo (2019), who use a multinomial logit demand system to infer investor preferences from portfolio holdings. Their central finding is that demand for stocks is far less price elastic than standard asset-pricing models suggest. The finding matters economically. If demand is this inelastic, modest shifts in flows or mandates can move prices substantially, which bears directly on important questions such as the transmission of policy through financial markets and market stability.

As the framework has spread, a growing literature has asked how its estimates should be interpreted. The debate draws on several features of the estimates. In unrestricted specifications, some coefficients on market capitalization imply upward-sloping demand; estimates are unstable. And most of the model's predictive power comes from the residual latent demand component rather than from the estimated preference parameters (van Binsbergen et al., 2026; Fuchs et al., 2025a). Alongside the low estimated elasticities, these findings have motivated a series of work questioning whether the estimated coefficients recover the structural objects of interest. That work has offered two broad explanations.

One explanation concerns the economic model itself. Asset demand may differ from product demand in ways the share system does not capture. Fuchs et al. (2025a) show that cross-asset substitution and equilibrium price spillovers can make a simple demand-system coefficient differ sharply from the underlying structural elasticity, and that conditioning on common characteristics can turn the estimated object into an elasticity of residual demand. Fuchs et al. (2025b) take the argument further, arguing that asset demand elasticities cannot be recovered from the data without strong assumptions on investor behavior. On this view, the difficulty lies in the economics of portfolio choice itself, which a share system with restricted substitution cannot fully represent.

A second explanation concerns identification within a maintained demand model. van Binsbergen et al. (2026) and He et al. (2026) show that in dynamic asset markets, the shocks used to identify demand can also change future investment opportunities, shifting the very demand curves they are intended to trace. Haddad et al. (2026) raise a complementary point: cross-sectional instruments identify a relative elasticity, while recovering the absolute elasticity requires time-series variation. On this view, the demand model may be adequate, but the cross-sectional variation used to estimate it identifies a different object from the structural elasticity.

There is, however, a third possibility: the demand model may be a useful approximation and the identification strategy may be valid, but the data contain some form of measurement error, so that the portfolio shares presented to the estimator differ from the allocations the logit model describes. This paper examines that possibility.

At first glance, the idea that portfolio shares can be noisily measured may sound implausible. Holdings data are often drawn from regulatory filings or administrative records intended to provide an accurate account of what investors own. The issue, however, is not that large reported positions are grossly mismeasured. It arises in the long tail of the portfolio, where the distinction between the allocation recorded in the data and the smooth share object in the model can become economically small but statistically important.

Portfolio shares sum to one, but institutional portfolios distribute that unit of capital across hundreds or thousands of securities. At one end, a handful of large stocks may each account for several percent of the portfolio. At the other, the portfolio contains a long tail of positions measured in basis points or fractions of a basis point. These positions carry little capital individually, and often little in aggregate, but they are precisely where omissions, stale holdings, underweighting, and reporting imperfections are most likely to occur. Form 13F, for example, explicitly permits managers to omit positions that satisfy both a share-count and market-value de minimis threshold; an index fund may also deliberately omit tiny benchmark constituents when it uses optimized sampling.

Why should economically immaterial positions matter for estimation? Because the conventional relative-error estimator (the ratio moment) does not weight them by their dollar importance. Under the maintained model, deviations are divided by the fitted share, so a position with a fitted weight of 0.01% is scaled by $1/\mu_j$ just as mechanically as a position with a weight of 1%. This normalization is efficient when both the signal and the disturbance shrink proportionately with position size: a smaller position carries less signal, but proportionately less noise as well. When measurement error does not shrink at the same rate, however, the normalization magnifies the error. A tail that represents only a small fraction of portfolio capital can therefore represent a large fraction of the estimator's effective observations.

It is perhaps noteworthy, then, that the problem is not that an estimator developed for share data is now intrinsically ill-suited for finance. It is that the distribution of those shares is fundamentally different. In a consumption goods market where consumers choose among twenty or so brands, for example, each brand represents an economically meaningful alternative (Nevo, 2001); an institutional portfolio can contain thousands, with much of that nominal cross section concentrated in economically negligible positions. The distinction is largely immaterial when the relative error model continues to hold into the tail. It becomes first-order when it does not.

In the empirical analysis, I show that seemingly minor departures in the portfolio tail can generate the central pathologies that have motivated the current debate. Index funds provide a particularly transparent diagnostic because their benchmark coefficient is known. The relative-error estimator systematically places the coefficient above the passive benchmark. The direction of the discrepancy is consistent with omission or underweighting of small stocks in the tail, which relative-error estimation reads as a preference for larger firms. The same upward distortion can push the coefficient beyond the downward-sloping-demand region.

Having isolated the source of the problem, the natural response is to change how deviations in the tail enter the estimating equation. I use a level-share moment that leaves discrepancies in absolute units rather than dividing them by fitted shares. In the simple exogenous case, this

moment has the same quasi-score structure as Poisson pseudo-maximum likelihood, connecting the approach to the literature on count and count-like outcomes in finance (Cohn et al., 2022). The economic implication is simple. If a small model position is omitted, its influence remains proportional to its portfolio share. Bias therefore scales with the capital carried by the problematic tail rather than with the number of names in it. In the index-fund tests, this estimator stays substantially closer to the passive benchmark than the relative-error estimator.

Measuring discrepancies in absolute units does not guarantee precise estimates for every fund. Precision depends on portfolio composition, instrument strength, and the structure of disturbances. I therefore report uncertainty at the fund level and account for it when combining estimated demand responses.

I apply the estimator to CRSP mutual-fund holdings using the same broad characteristic specification as the canonical implementation. Estimated demand coefficients are economically interpretable and stable over time without imposing sign restrictions. Characteristic coefficients distinguish recognizable investment strategies, including growth and value mandates, while stock-level demand is almost uniformly downward-sloping. Estimated mutual-fund demand is substantially more elastic than relative-error estimation implies, although it remains inelastic. Its elasticity declines over the sample as passive ownership expands and demand among non-passive holders also becomes less elastic.

To evaluate the economic content of these estimates, I construct an out-of-sample structural test of funds' trading responses to price changes. The test uses coefficients estimated two quarters earlier and compares funds facing the same prices. Under the model's substitution structure, the design absorbs the effects of other price changes, separating investors' demand responses from general-equilibrium price adjustments. With price changes instrumented by flow-induced trading, funds implement about two-thirds of the predicted differences in trading within a quarter. A simple demand system estimated from portfolio allocations therefore captures economically meaningful differences in how investors respond to prices.

The contribution is primarily empirical and econometric. The analysis complements work on substitution and dynamic identification by examining how discrepancies in portfolio holdings affect estimation and by evaluating the resulting demand curves against subsequent behavior. Several empirical features that have motivated doubts about demand systems prove sensitive to how holdings enter estimation. Accounting for that sensitivity produces more stable estimates with predictive economic content, providing an empirical foundation for further development of the framework.

Related literature. *Demand-system asset pricing.* This paper builds on Koijen and Yogo (2019) and Koijen, Richmond and Yogo (2024), who develop and apply a structural demand approach to asset pricing using investor portfolio holdings. Subsequent work has extended the framework across asset classes and economic settings,[1] while also advancing the empirical methodology. Graves (2024), for example, develops an approach for censored institutional holdings and heterogeneous portfolio constraints, while van der Beck (2026) introduces portfolio dy-

[1] Applications include euro-area government bonds (Koijen et al., 2021), corporate bonds (Bretscher et al., 2026), international portfolios spanning equities, bonds, and currencies (Koijen and Yogo, 2026), and Treasury markets (Jansen et al., 2026).

namics to distinguish short- from long-run demand elasticities. Haddad et al. (2025) incorporate strategic interaction among investors into the demand system and show that investors' strategic responses offset only part of a change in others' behavior, so that the rise of passive investing has made stock-level demand more inelastic. This paper contributes along a complementary empirical margin. Rather than extending the model structure of the demand system, I focus on the estimators and the related inference procedure: how demand estimation can better accommodate the concentration and measurement-error structure of financial holdings.[2]

Inelastic demand and substitution. A long empirical literature shows that demand curves for individual stocks slope down. Early evidence comes from index additions and weight changes (Shleifer, 1986; Harris and Gurel, 1986; Kaul et al., 2000), and later work links the size of these price effects to the availability of close substitutes (Wurgler and Zhuravskaya, 2002). Regression-discontinuity and flow-based designs reach similar conclusions (Chang et al., 2015; Lou, 2012), and Gabaix and Koijen (2021) argue that the aggregate market is itself inelastic. These estimates are orders of magnitude below the elasticities implied by classic models (Petajisto, 2009), which has prompted work on why demand is inelastic and how substitution shapes the answer. Davis et al. (2025) show that standard portfolio-choice models imply highly elastic demand only when price changes pass through fully to expected returns and assets have close substitutes; relaxing either assumption delivers elasticities near empirical estimates. Chaudhary et al. (2026) show that allowing heterogeneous substitution among corporate bonds makes security-level multipliers an order of magnitude smaller, while aggregated portfolios retain substantially larger multipliers. Li and Lin (2026) show that price multipliers increase with the systematic risk carried by order flows, connecting substitution and risk-bearing capacity to differences in price impact. Holdings-based demand estimates are often compared with these price-impact designs. This paper adds evidence from the trading side: whether investors' estimated elasticities predict how they trade when prices move.

Measurement, transformation, and aggregation. This paper is closest in spirit to a set of empirical studies showing that seemingly innocuous choices about how economic objects are measured, transformed, or constructed can have material impact on what an estimator recovers. Erickson and Whited (2000) show that measurement error in q changes the interpretation of investment regressions; Cohn et al. (2022) show that transformations of count and count-like outcomes can change the estimand and even the expected sign of a coefficient; and Gonçalves et al. (2020) show that production-based asset-pricing estimates depend on whether returns are constructed before or after aggregation. Although the present paper studies a different economic model and setting, it owes much of its empirical perspective to the insights and inspiration of this literature. Relative-error normalization determines which positions receive statistical leverage, while consolidation determines which allocation is treated as a single decision. The broader lesson is that seemingly technical choices made before estimation—choices that may appear peripheral to the economic model itself—can have first-order effects on the economic object recovered from the data and, ultimately, on the conclusions drawn from it.

[2]The level formulation relates to quasi-likelihood methods for fractional and multivariate share responses (Papke and Wooldridge, 1996; Mullahy, 2015), to the equivalence between conditional logit and Poisson estimation (Guimarães et al., 2003), and to the observation that the appropriate estimating equation depends on how the conditional variance scales with the conditional mean (Santos Silva and Tenreyro, 2006).

Outline. Section 1 describes the holdings data and where they depart from what a demand model implies. Section 2 sets out a simple allocation model and the two estimators, shows how each responds to measurement error in the tail, and compares them on index funds, whose value-weighting benchmark is known. Section 3 describes exactly what is estimated in the full data and how, shows what the estimates look like, and compares them with those of the ratio moment. Section 4 tests the estimates out of sample, in the funds' subsequent trading. Section 5 concludes.

1 Portfolio Data and Measurement Error

Institutional holdings provide a detailed and accurate record of portfolio allocations. Their use in demand estimation requires an additional step: interpreting those allocations as the outcome of investor preferences at prevailing prices. Reporting rules and portfolio implementation can create a gap between the two. Such gaps can be small enough to have little effect on the portfolio while materially affecting the demand inferred from it. This section documents the portfolio structure that makes this possible and the sources of the gap.

1.1 Concentration and the portfolio tail

Financial markets are well known to be highly concentrated. It is also elementary that portfolio weights are nonnegative and sum to one.[3] What is less appreciated is how these two facts come together in portfolio data. When a portfolio spans hundreds or thousands of securities but a small number of large positions absorb a substantial share of its capital, the remaining capital must be spread thinly across the rest. Concentration at the top therefore mechanically produces a long tail of very small portfolio weights, often measured in basis points or fractions of a basis point.

The economic scale of these positions is easy to overlook. A holding of 0.01 percent represents just one basis point of the portfolio, yet institutional portfolios can contain hundreds or thousands of positions at this scale. Such holdings can constitute much of the observed cross section while accounting for little of the invested capital.

[3]Other data in finance share this structure, among them bank–firm lending shares (Khwaja and Mian, 2008; Chodorow-Reich, 2014; Amiti and Weinstein, 2018), syndicate participations (Sufi, 2007) and index weights. The empirical analysis here uses equity holdings.

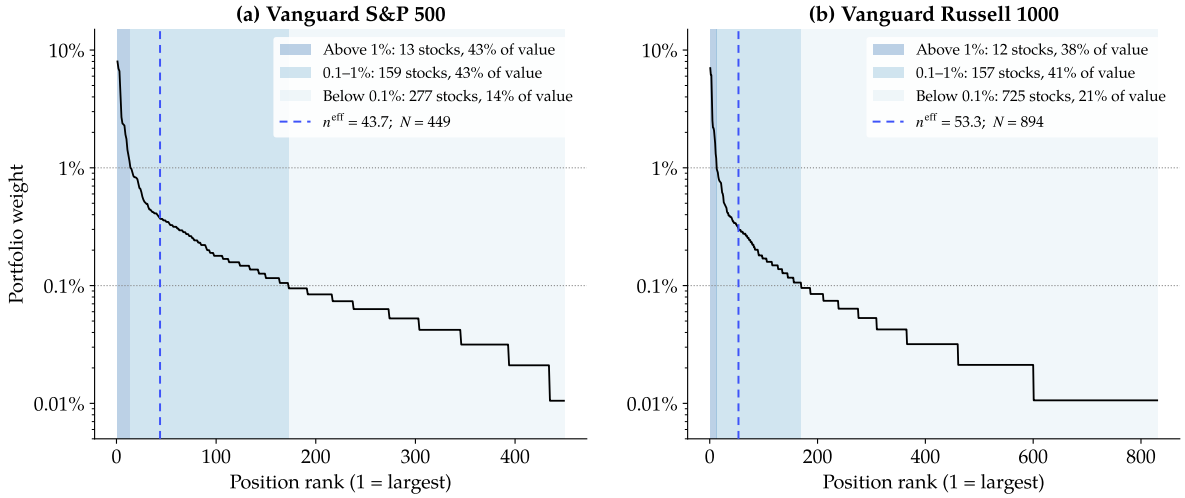

Figure 1: **Portfolio weights of two index funds have a long tail of very small positions.** The panels plot the portfolio weights of the Vanguard S&P 500 fund (a) and the Vanguard Russell 1000 fund (b), whose ETF share classes trade as VOO and VONE, in 2024Q4, sorted by size (rank 1 = largest) on a log scale with a common vertical axis. Shaded bands mark positions above 1%, between 0.1% and 1%, and below 0.1% of the portfolio; the legend gives the number of stocks and the share of value in each band, and the dashed line marks the inverse Herfindahl index $n^{\text{eff}} \equiv 1/\sum_j w_j^2$, a measure of portfolio concentration, alongside N, the number of positions held. Weights are CRSP reported weights (`percent_tna`), which are rounded to one basis point, so positions below about 0.005% of the portfolio are reported as zero; they are counted in N and the lowest band but cannot be drawn.

Figure 1 illustrates this pattern for two large index funds in 2024Q4, the Vanguard S&P 500 and Vanguard Russell 1000 funds. In the S&P 500 fund, 13 positions exceeding 1 percent of the portfolio account for 43 percent of its value, while 277 positions below 0.1 percent account for just 14 percent. The contrast becomes sharper as the investment universe expands: in the Russell 1000 fund, 151 of its 894 positions carry less than one basis point each. Expanding the portfolio adds many securities while allocating little capital to each additional name. The pattern extends beyond index funds. Across the mutual-fund sector in 2024, positions below 0.1 percent account for 31 percent of holdings by count but only about 7 percent of portfolio value. A related distinction is familiar from empirical asset pricing: Fama and French (2008) observe that microcaps constitute roughly 60 percent of stocks in their sample but only 3 percent of aggregate market capitalization. Here the disparity arises within an investor's portfolio, and it can be substantial even in a universe restricted to large-cap stocks, as the S&P 500 fund shows. It also separates portfolio data from the product markets in which demand systems were first developed. A consumer choosing among twenty or so brands of breakfast cereal faces alternatives that each hold a meaningful share of the market (Nevo, 2001); most of the cross section of an institutional portfolio consists of positions that hold almost none of its capital.

1.2 Reported holdings and target allocations

Small positions allow considerable latitude in portfolio implementation. A fund can omit a holding targeted at one basis point, or hold twice that amount, while changing its overall allocation by only one basis point. Comparable relative departures in a large position would substantially alter the portfolio. Investors can therefore have little incentive to implement

every small position precisely, even when their desired allocation is well defined.

Index funds provide a clear example. Representative or optimized sampling allows a fund to track its benchmark while holding only a subset of its constituents. The fund trades off the benefits of closer replication against the costs of maintaining additional positions. Its holdings can consequently depart from benchmark weights as part of its implementation strategy. Fund prospectuses explicitly allow such omissions (BlackRock, 2025), and Brogaard et al. (2026) document that sampling systematically underweights or omits illiquid constituents.[4]

Trading costs create a related gap as portfolios adjust over time. Investors need not immediately rebalance every position when prices, characteristics or desired allocations change. With costly adjustment, current holdings can partly reflect earlier allocation decisions (Gârleanu and Pedersen, 2013). Relative to a static demand model, these holdings depart from the allocation implied by current conditions.[5] Reporting conventions can introduce further discrepancies: weights may be rounded, and sufficiently small positions may be omitted under applicable reporting rules.[6]

These mechanisms need not produce discrepancies that shrink proportionately with position size. Reporting thresholds operate in shares or dollars, while portfolio implementation balances the benefit of closer tracking or rebalancing against its cost. Consider a stock targeted at 10 percent of the portfolio but held at 10.01 percent. The departure is 0.01 percentage points, or 0.1 percent of the target. Omitting a stock targeted at 0.01 percent produces the same absolute departure but a 100 percent shortfall relative to its target. Both departures involve one basis point of portfolio capital; their relative magnitudes differ by a factor of one thousand.

The distinction is between the allocation described by the demand model and the holding observed by the researcher. Let w_j^* denote the target weight of security j, incorporating the investor's preferences and beliefs at prevailing prices, and let $\tilde{w}_j$ denote its observed weight. Then

$$\tilde{w}_j = w_j^* + c_j, \tag{1}$$

where c_j collects implementation and reporting departures in absolute units, as shares of the portfolio. Following Erickson and Whited (2000), I use measurement error to describe the gap between the object specified by the economic model and its observed counterpart. A filing can accurately record what a fund owns while differing from the allocation the demand model describes. The resulting discrepancy can be systematic, as when sampling excludes small constituents, and can remain substantial relative to the target even when it involves little portfolio capital. For estimation, the relevant feature of these departures is how their magnitude and direction vary with position size. Section 2 shows why, and what it implies for the choice of estimator.

[4]Related evidence comes from corporate bond markets, where Li et al. (2026) find that fund portfolios are sparse relative to their benchmarks and that benchmark weights predict which bonds funds hold. They model this sparsity as the choice of benchmarked managers who face portfolio-management costs.

[5]This source of departures could instead be modeled as part of the investor's portfolio dynamics (van der Beck, 2026).

[6]Form 13F, for example, permits managers to omit positions of fewer than 10,000 shares and under $200,000 in market value (U.S. Securities and Exchange Commission, 2025). Both conditions must hold, and omission is permitted rather than required.

2 Estimating Demand from Portfolio Holdings

Section 1 described how reporting and implementation can separate observed holdings from target allocations, particularly among small positions. This section examines how those departures affect demand estimation. I begin with the characteristics-based demand system of Koijen and Yogo (2019) and the identifying restriction underlying its standard estimator. I then examine how measurement error enters that restriction and develop an alternative based on discrepancies measured in absolute units. Index funds provide a benchmark for comparing the resulting estimates, while the final subsection examines why pooling portfolios, the conventional response to imprecise estimates, does not simply enlarge the sample.

2.1 Demand and the standard identifying restriction

Consider one investor allocating capital across an investment universe U in a given quarter. I suppress investor and time subscripts to emphasize that estimation uses the cross section of securities within a portfolio. Following Koijen and Yogo (2019), target demand depends on market capitalization, observed stock characteristics, and latent demand:

$$w_j^* = \mu_j(\theta)\varepsilon_j, \qquad \mu_j(\theta) = \exp\left(a + \alpha \log \mathrm{ME}_j + X_j'\beta\right). \tag{2}$$

Here, ME_j is market capitalization, X_j contains observed characteristics, and a is a common portfolio intercept that absorbs the budget normalization. The parameter vector θ collects the intercept and demand coefficients. Latent demand, $\varepsilon_j \geq 0$, captures the investor's preference for security j beyond what the observed characteristics explain.

The multiplicative specification expresses latent demand as a relative tilt to characteristics-based demand. A given relative tilt represents a larger allocation departure in a large position than in a small one. This form follows from allowing unobserved characteristics to enter the exponential demand index alongside observed characteristics. The distribution of latent demand is otherwise left unspecified.

The coefficient α measures how portfolio weights vary with market capitalization, holding the other characteristics fixed. A capitalization-weighted portfolio is reproduced by $\alpha = 1$, zero coefficients on the other characteristics, and constant latent demand. This provides a benchmark for the index-fund comparisons below.

Recovering the demand coefficients requires a restriction on latent demand. Market capitalization contains price, which responds to investors' demand. Securities with high latent demand can therefore have high prices even when an individual investor has negligible price impact, because the investor's demand shocks may be correlated with those of other investors. Cross-sectional variation in market capitalization alone need not identify the investor's response to price.

Koijen and Yogo (2019) address this endogeneity using variation in other investors' investment mandates and capital. Their instrument assigns each other investor's assets under management equally across the securities in its investment universe and aggregates the resulting allocations to each stock. A stock eligible for more investment capital receives a higher

instrument value. The investor whose demand is being estimated is excluded from the construction.

The economic argument is that other investors' predetermined mandates and capital shift demand for a stock and hence its price. Conditional on observed characteristics, this variation is assumed to be unrelated to the focal investor's latent demand. Let z_j collect the excluded instrument and the included characteristics, together with a constant. The identifying restriction is

$$E[\varepsilon_j \mid z_j] = 1. \tag{3}$$

The unit mean fixes the scale of latent demand relative to the intercept. The identifying content is that its conditional mean does not vary with z_j. Thus, stocks exposed to more mandate-driven capital may have higher prices, but they do not systematically receive larger or smaller latent-demand tilts from the investor being estimated.

Because $\varepsilon_j = w_j^* / \mu_j(\theta_0)$, the restriction implies

$$E\left[z_j \left\{ \frac{w_j^*}{\mu_j(\theta_0)} - 1 \right\}\right] = 0. \tag{4}$$

Replacing expectations with cross-sectional averages and target weights with observed holdings gives the standard estimating equation:

$$\sum_{j \in U} z_j \left\{ \frac{\tilde{w}_j}{\mu_j(\hat{\theta})} - 1 \right\} = 0. \tag{5}$$

I refer to this as the ratio moment. It measures each discrepancy relative to the fitted holding. Applying it to observed holdings requires examining how the reporting and implementation departures described in Section 1 affect the restriction.

2.2 Measurement error and the ratio moment

The ratio moment has a natural rationale when disturbances are proportional to position size. A small holding then contains less signal about demand, but its disturbance is correspondingly smaller. Dividing by the fitted weight puts discrepancies on a common relative scale, allowing small positions to contribute information alongside large ones. The usefulness of this normalization depends on whether the departures in observed holdings follow the same scaling.

Section 1 distinguished target allocations from observed holdings. Combining that distinction with the demand specification gives

$$\tilde{w}_j = \mu_j(\theta_0)\varepsilon_j + c_j. \tag{6}$$

Evaluated at the true parameter, the ratio residual becomes

$$\frac{\tilde{w}_j}{\mu_j(\theta_0)} - 1 = \underbrace{\varepsilon_j - 1}_{\text{Latent demand}} + \underbrace{\frac{c_j}{\mu_j(\theta_0)}}_{\text{Measurement error}} . \tag{7}$$

The normalization removes position size from the latent-demand component and places its inverse on measurement error. Consequently, a discrepancy that is small in absolute units can remain substantial relative to a small fitted holding.

This changes what is required for the estimating equation to recover the demand parameters. Under the identifying restriction in Section 2.1,

$$E\left[z_j \left\{ \frac{\tilde{w}_j}{\mu_j(\theta_0)} - 1 \right\}\right] = E\left[z_j \frac{c_j}{\mu_j(\theta_0)}\right] . \tag{8}$$

Applying the ratio moment to observed holdings therefore requires an additional orthogonality restriction on measurement error relative to the fitted position. Orthogonality of the absolute discrepancy, $E[z_j c_j] = 0$, does not generally imply this restriction, because the fitted weight contains endogenous prices.

Even when this additional contribution is mean zero, it can make small positions substantially noisier than the proportional-disturbance model suggests. When it varies systematically with the instrument, it can also shift the estimated coefficients.

Omission illustrates the issue directly. An observed zero produces a ratio residual of -1, regardless of the fitted position's size. A portfolio can therefore contain many complete relative shortfalls while allocating little capital to the affected securities. Their contribution to estimation depends on the associated instrument values. If omitted or underweighted positions are systematically small stocks, the resulting pattern can be interpreted as a preference for larger stocks.

2.2.1 An index-fund benchmark

Broad index funds have observable reference allocations against which estimated demand can be assessed. For a capitalization-weighted portfolio, the demand specification in equation (2) reproduces the reference allocation at $\alpha = 1$, with zero coefficients on the other characteristics and constant latent demand. Sampling, trading, and reporting can move observed holdings away from these allocations without representing a different underlying investment mandate. These funds therefore provide a useful benchmark for examining whether estimation attributes implementation departures to demand preferences.

I examine two Vanguard funds tracking the S&P 500 and Russell 1000. The sample runs quarterly from 2010Q2 to 2024Q4, with the Russell 1000 fund entering in 2010Q3. Estimation uses the specification and instrument employed in the full demand system described in Section 3.

Figure 2 plots fund-quarter estimates for the two index funds. Each marker represents an estimate for one fund in one quarter, and the dashed vertical line marks the capitalization-weighting benchmark of one. For each security, estimated latent demand is $\hat{\varepsilon}_j = \tilde{w}_j / \hat{\mu}_j$. The

vertical axis reports its portfolio-weighted mean in logs. The black crosses report ratio-moment estimates. For both funds, these estimates lie systematically to the right of the benchmark. The estimated coefficient exceeds one in every quarter, with mean estimates of 1.10 and 1.13.

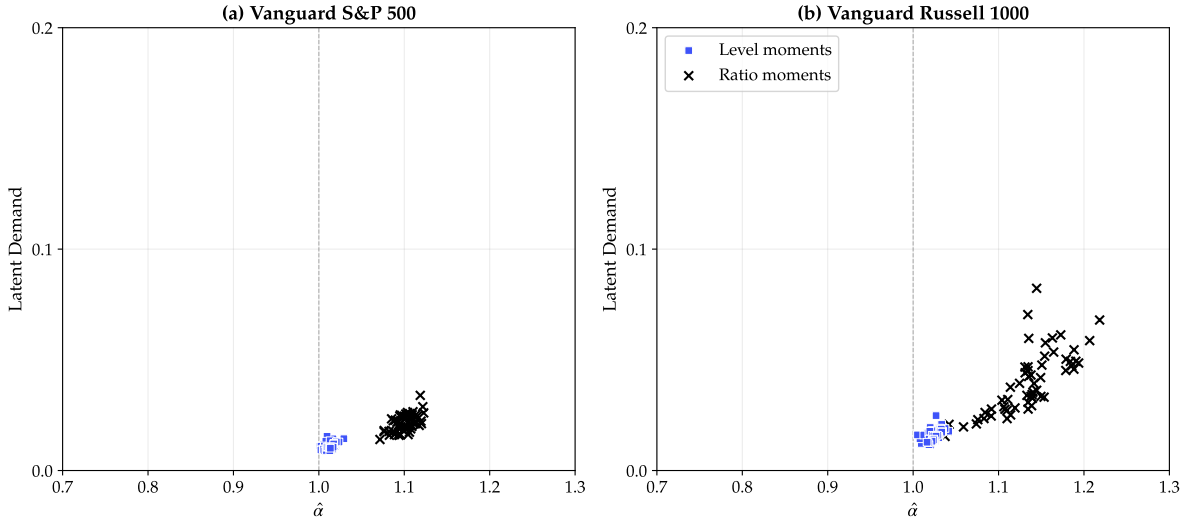

Figure 2: **Level and ratio estimates for two Vanguard index funds.** The panels plot fund-quarter estimates for the Vanguard S&P 500 fund (a), 2010Q2–2024Q4, and the Vanguard Russell 1000 fund (b), 2010Q3–2024Q4, whose ETF share classes trade as VOO and VONE. Holdings are those of each fund as a whole, common to its mutual-fund and ETF share classes. The horizontal axis is the price coefficient $\hat{\alpha}$ of the baseline demand specification (Section 3.1) recovered by each moment, and the vertical axis is the portfolio-weighted mean of estimated log latent demand, $\sum_{j:\, w_j>0} w_j \log(w_j/\hat{\mu}_j)$, with $\hat{\mu}_j$ the fitted share. Stocks in the fund's universe that it does not hold have zero weight and drop out of the mean; their log latent demand is undefined. Blue squares are level-moment estimates and black crosses are ratio-moment estimates; the dashed vertical line marks the capitalization-weighting benchmark of one.

The departure from one has a direct economic interpretation. When a constituent's price rises, a capitalization-weighted fund's existing holding appreciates with it. The fund need not buy additional shares to accommodate that increase in the stock's capitalization. A coefficient above one instead attributes a more-than-proportional increase in the desired dollar allocation to the price increase, and therefore places these passive funds in the upward-sloping region of the demand model.

This interpretation makes the benchmark informative. The estimator is applied to funds with explicitly capitalization-weighted mandates, yet recovers coefficients suggesting that they buy additional shares as prices rise. The direction of the discrepancy is consistent with the mechanism in Section 2.2: relative shortfalls among small constituents can be interpreted as a stronger preference for market capitalization. Establishing how much of the discrepancy comes specifically from omission requires measuring those departures directly. The estimates nevertheless show that the choice of moment can materially affect the demand behavior attributed to a known investment strategy.

The figure also reports level-moment estimates, shown as blue squares and developed in the next subsection. These are substantially closer to the benchmark, with mean coefficients of 1.01 and 1.03. The remaining departure indicates that changing the moment does not eliminate every discrepancy between observed holdings and reference allocations.

2.3 Estimation in absolute units

I estimate demand using discrepancies measured in absolute units. The level estimator solves

$$\sum_{j \in U} z_j \{ \tilde{w}_j - \mu_j(\hat{\theta}) \} = 0. \tag{9}$$

This equation matches the observed portfolio's exposure to the instruments and included characteristics with the exposure implied by the fitted portfolio.[7] For an included characteristic, the corresponding equation matches its portfolio-weighted average. With a constant among the instruments, the fitted weights also sum to the observed portfolio total. The estimator therefore permits discrepancies across individual securities while requiring their instrument-weighted contributions to balance at the portfolio level. An underweight in one security can be offset by an overweight elsewhere, with the offset depending on the securities' characteristics and instrument values.

The distinction from ratio estimation is how each allocation discrepancy enters the equation. Under the ratio moment, an omitted position contributes $-z_j$, regardless of its fitted weight. Under the level moment, it contributes $-z_j \mu_j(\theta)$. Omitting a small fitted position therefore has a correspondingly small direct effect on the moment. The same principle applies to underweights and overweights: their contributions depend on the portfolio capital involved.

This treatment is useful for the sampling decisions described in Section 1. Omitting a group of small constituents creates underweights in those securities and reallocates their capital elsewhere. The level moment records both sides of that reallocation. Its response depends on how much capital moves and how the instruments differ between the underweighted and overweighted securities. A large number of omissions can consequently have a modest effect when little capital is involved.

The corresponding population restriction is

$$E\left[z_j \{ \tilde{w}_j - \mu_j(\theta_0) \} \right] = 0. \tag{10}$$

It requires the total allocation discrepancy, measured in absolute units, to be orthogonal to the instruments and included characteristics. Because fitted shares contain endogenous prices, this restriction does not follow automatically from the relative-error restriction (3): weighting relative latent-demand departures by fitted position size can reintroduce dependence on the instruments through price. Appendix A.2 examines this restriction and its sensitivity to systematic departures. It also reports an estimator that weights relative discrepancies by instrument-predicted rather than fitted positions, which retains the relative-error restriction on latent demand. Its estimates are nearly identical to the level estimates: the asset-weighted mean price coefficient is 0.710, against 0.707 under the level moment and 0.811 under the ratio

[7] When the characteristics serve as their own instruments, the level moment is the Poisson pseudo-maximum-likelihood score and the ratio moment the Gamma score; the choice between them is a choice of variance function (Santos Silva and Tenreyro, 2006). Appendix A.1 gives the mapping, relates it to fractional-response estimators (Papke and Wooldridge, 1996; Mullahy, 2015), and distinguishes it from the levels-versus-logs comparison in Koijen and Yogo (2019).

moment.

Returning to the two index funds examined in Section 2.2.1, the level estimates are substantially closer to the capitalization-weighting benchmark (Figure 2). Measuring discrepancies in absolute units therefore brings the estimated allocation rule much closer to the funds' stated mandates, which illustrates the practical importance of how allocation discrepancies enter estimation.

2.4 The conventional response: pooling portfolios

A familiar response to measurement error is to increase sample size: additional observations allow idiosyncratic disturbances to average out and improve the precision of estimated coefficients. Existing demand-system applications follow this logic by pooling observations across similar investors when individual portfolios provide insufficient information for accurate estimation or estimation encounters convergence difficulties. Pooling therefore offers a natural response to the estimation problems discussed above.

For portfolio demand, however, combining holdings does more than increase the number of observations. It also combines allocation rules. Because the logit demand system is nonlinear, adding portfolios generated by different coefficients generally does not produce a portfolio that follows the same specification with average coefficients. The potential gain from averaging out measurement noise therefore comes with a change in the demand relationship being estimated.

Appendix B expresses the consolidated demand index as

$$\Lambda_j = \log\left(\sum_m \pi_m e^{\delta_{jm}}\right) = \log E_\pi\left[e^{\delta_{jm}}\right], \qquad \delta_{jm} = \alpha_m \log \mathrm{ME}_j + X_j'\beta_m, \tag{11}$$

where π_m are the normalized mixing weights defined there. This is the cumulant generating function of the constituent demand indices, evaluated at one. Its expansion begins with

$$\Lambda_j = E_\pi\left[\delta_{jm}\right] + \frac{1}{2}\mathrm{Var}_\pi\left(\delta_{jm}\right) + \cdots. \tag{12}$$

The first term is an average demand index. The remaining terms reflect heterogeneity across mandates. Consolidation therefore brings the dispersion of allocation rules into the demand object presented to the estimator.

The resulting curvature has a simple interpretation. Holding the mixing weights fixed, a higher market capitalization gives greater relative influence to mandates with larger capitalization coefficients. The slope of consolidated demand consequently changes with market capitalization, even though each constituent index has a constant slope. In this representation, curvature in the price direction equals the variance of those coefficients under the stock-specific mixing weights.

Fitting a single linear demand index to the consolidated portfolio approximates this curved relationship with one slope. That coefficient depends on the securities observed, their characteristics, and the estimating moments; it need not equal the capital-weighted average of the constituent coefficients. The component left outside the fitted index enters the residual as ap-

parent latent demand. Heterogeneity thus creates a systematic specification component rather than another disturbance that necessarily disappears with more observations. Appendix B develops the exact representation and the role of portfolio normalization.

Consolidating holdings before estimation and combining estimated demand responses afterward are different operations. The former fits a common demand rule to the combined portfolio, whose allocation need not follow the specification of its constituent mandates. The latter estimates each mandate's allocation rule separately and then combines the estimated responses using the weights appropriate to the economic object. This preserves heterogeneity across mandates while allowing their estimation uncertainty to be carried into the combined response.

The implication for measurement error is that increasing the sample cannot substitute for addressing how discrepancies enter the estimating equation. Systematic implementation departures can remain in a larger sample, while consolidation can add discrepancies arising from heterogeneous demand rules. I therefore address measurement error within each portfolio through the level moment developed above, and combine estimated demand responses afterward when the application requires it.

3 Data and Estimated Demand

Section 2 developed the level moment for estimating demand from individual portfolios. I now apply it to a panel of U.S. equity mutual funds. Section 3.1 describes the data, demand specification, and estimation procedure. Section 3.2 presents the estimated demand coefficients, and Section 3.3 compares them with estimates obtained from the ratio moment.

3.1 Data, demand specification, and estimation

Holdings come from the CRSP Mutual Fund Database. The unit of observation is a fund portfolio in a quarter. Share classes investing in the same underlying portfolio are combined, while distinct funds within a family remain separate. A fund portfolio more closely corresponds to a single allocation decision than a 13F institutional filing, which can combine several mandates with different objectives. Section 2.4 explains why this distinction matters for demand estimation.

The sample covers funds classified by Lipper as equity funds and their holdings of U.S. common stocks from 2010Q2 to 2024Q4. Following Koijen and Yogo (2019), I construct a rolling investment universe: $\mathcal{U}_{ft}$ contains every stock that fund f holds in quarter t or held in any of the preceding eleven quarters. Observed portfolio weights, denoted w_{fjt} throughout this section, are computed from position market values and normalized to sum to one over $\mathcal{U}_{ft}$. The estimated demand system therefore describes allocation within the U.S. equity universe.

For each fund-quarter (f,t) and stock $j \in \mathcal{U}_{ft}$, the systematic demand index is

$$\eta_{fjt}(\theta_{ft}) = \alpha_{ft} \log \mathrm{ME}_{jt} + \beta_{ft}^{BM} \log(\mathrm{B/M})_{jt} + X_{jt}'\beta_{ft}, \tag{13}$$

with fitted portfolio shares

$$\mu_{fjt}(\theta_{ft}) = \frac{\exp\{\eta_{fjt}(\theta_{ft})\}}{\sum_{k\in\mathcal{U}_{ft}} \exp\{\eta_{fkt}(\theta_{ft})\}}. \tag{14}$$

The vector X_{jt} contains operating profitability, investment, market beta, and dividend characteristics. Their construction follows Koijen and Yogo (2019), with a six-month lag for accounting information. Book-to-market enters in place of Koijen and Yogo (2019)'s log book equity; given log market equity the two are equivalent (Appendix C.3), and book-to-market separates the value–growth taste from the price slope. Because portfolio shares are normalized within the equity universe, a common intercept cancels from the specification. Adding 12–2 month momentum to X_{jt} leaves the price coefficients essentially unchanged (Appendix C.4); the trading test in Section 4 uses the funds' momentum tastes from that version.

I instrument log market equity with a leave-one-out demand shifter constructed from other funds' assets under management and investment universes. The remaining characteristics serve as their own instruments. In implementation, log book equity takes the place of log book-to-market: given log market equity the two specifications are equivalent, and log book equity, which does not move with the price, serves as its own instrument, whereas log book-to-market would inherit the endogeneity of the price. The estimates are converted to the book-to-market basis (Appendix C.3). The level moment from Section 2 becomes

$$g_{ft}(\theta_{ft}) \equiv \sum_{j\in\mathcal{U}_{ft}} z_{fjt}\{w_{fjt} - \mu_{fjt}(\theta_{ft})\} = 0, \tag{15}$$

where z_{fjt} contains the excluded instrument and the included characteristics. I solve this just-identified system separately for each fund-quarter, without sign restrictions, cross-quarter pooling, or shrinkage. Each coefficient vector therefore describes the allocation rule of one fund at one date. Appendix C provides the details.

The coefficient α_{ft} measures the response of portfolio weights to market capitalization, holding book-to-market and the other characteristics fixed. Under this conditioning, it is the fund's price slope. A capitalization-weighted portfolio has $\alpha_{ft} = 1$ and zero coefficients on the remaining characteristics. The corresponding conditional quantity-demand elasticity for fund f and stock j is

$$\kappa_{fjt} = 1 - \alpha_{ft}(1 - w_{fjt}), \qquad \kappa_{jt} = \sum_f \omega_{fjt}\,\kappa_{fjt}, \tag{16}$$

where ω_{fjt} is fund f's share of the stock's mutual-fund ownership. Thus, κ_{jt} is the ownership-weighted elasticity of mutual-fund demand for stock j. A capitalization-weighted fund has $\kappa_{fjt} = w_{fjt}$, which is close to zero; lower values of α_{ft} imply more elastic demand for a given portfolio weight. Appendix C.2 develops the coefficient interpretation and distinguishes this conditional response from a price change holding book equity fixed.

Estimation begins with 264,766 fund-quarters. Of these, 7.62%, representing 3.94% of assets under management, have no moment root or an undefined sampling variance. I require a first-stage F-statistic of at least 10 and an effective number of positions of at least 10, and exclude a small number of additional estimation failures and price-weighted funds. The final

sample contains 212,997 fund-quarters, covering 80.4% of the initial observations and 89.9% of their assets under management. For aggregate statistics, I winsorize the estimated coefficients within each quarter at the 1st and 99th percentiles. Appendix C.3 reports the diagnostics and sample accounting.

3.2 Estimates from U.S. equity mutual funds

The estimates are built in two steps. Section 3.2.1 shows the fund-level coefficients for one fund family. Section 3.2.2 then sums the fund–stock elasticities across each stock's holders to obtain the stock-level elasticity.

3.2.1 Fund-level estimates

I first examine the 30 largest BlackRock and iShares funds in the 2024Q4 sample, which account for 85% of the family's assets in the sample. These funds span broad market indexes, size and sector indexes, growth and value strategies, and other investment mandates. They provide a benchmark for assessing whether the estimated coefficients distinguish recognizable allocation rules.

For presentation, I divide the funds into five groups: broad index, band or sector index, growth, value, and other active funds. The classifications use point-in-time Lipper classes, CRSP index flags, and fund names. They are used only to interpret the estimates; every fund is estimated using the same specification and procedure.

Figure 3 reports each fund's 2024Q4 price coefficient, $\hat{\alpha}$, and book-to-market coefficient, $\hat{\beta}^{BM}$. Markers show the quarter-end estimates, while lines span the minimum and maximum estimates over 2024Q1–2024Q4. The median within-year range for $\hat{\alpha}$ is 0.04. These ranges describe variation across quarterly estimates and are not confidence intervals.

Broad index funds lie close to the capitalization-weighting benchmark of $\alpha = 1$ and $\beta^{BM} = 0$. All six have $\hat{\alpha}$ between 1.01 and 1.04 and $\hat{\beta}^{BM}$ within 0.03 of zero, with estimates moving by less than 0.01 during the year. The iShares Core S&P 500 ETF and the iShares S&P 500 Index Fund are separate portfolios tracking the same index; both have $\hat{\alpha} = 1.013$. Size-band index funds have lower price coefficients, ranging from 0.58 for the Russell 2000 fund to 0.83 for the S&P small-cap fund. Unlike broad market trackers, these funds weight stocks by capitalization within a restricted segment of the market, while their rolling estimation universes can extend beyond that segment.

The book-to-market coefficients distinguish growth from value strategies. All five growth funds have negative $\hat{\beta}^{BM}$, ranging from -0.57 to -0.25, while all four value funds have positive coefficients, ranging from 0.05 to 0.45. Among the other active funds, two dividend funds and a minimum-volatility fund have the lowest price coefficients, with $\hat{\alpha}$ between 0.22 and 0.35. These estimates are consistent with strategies that depart substantially from capitalization weighting.

Several growth and factor strategies have price coefficients above one. The Russell 1000 Growth and Russell Top 200 Growth funds have $\hat{\alpha} = 1.16$ and 1.21, respectively, while an actively managed factor-rotation fund has $\hat{\alpha} = 1.29$. The technology fund's coefficient also

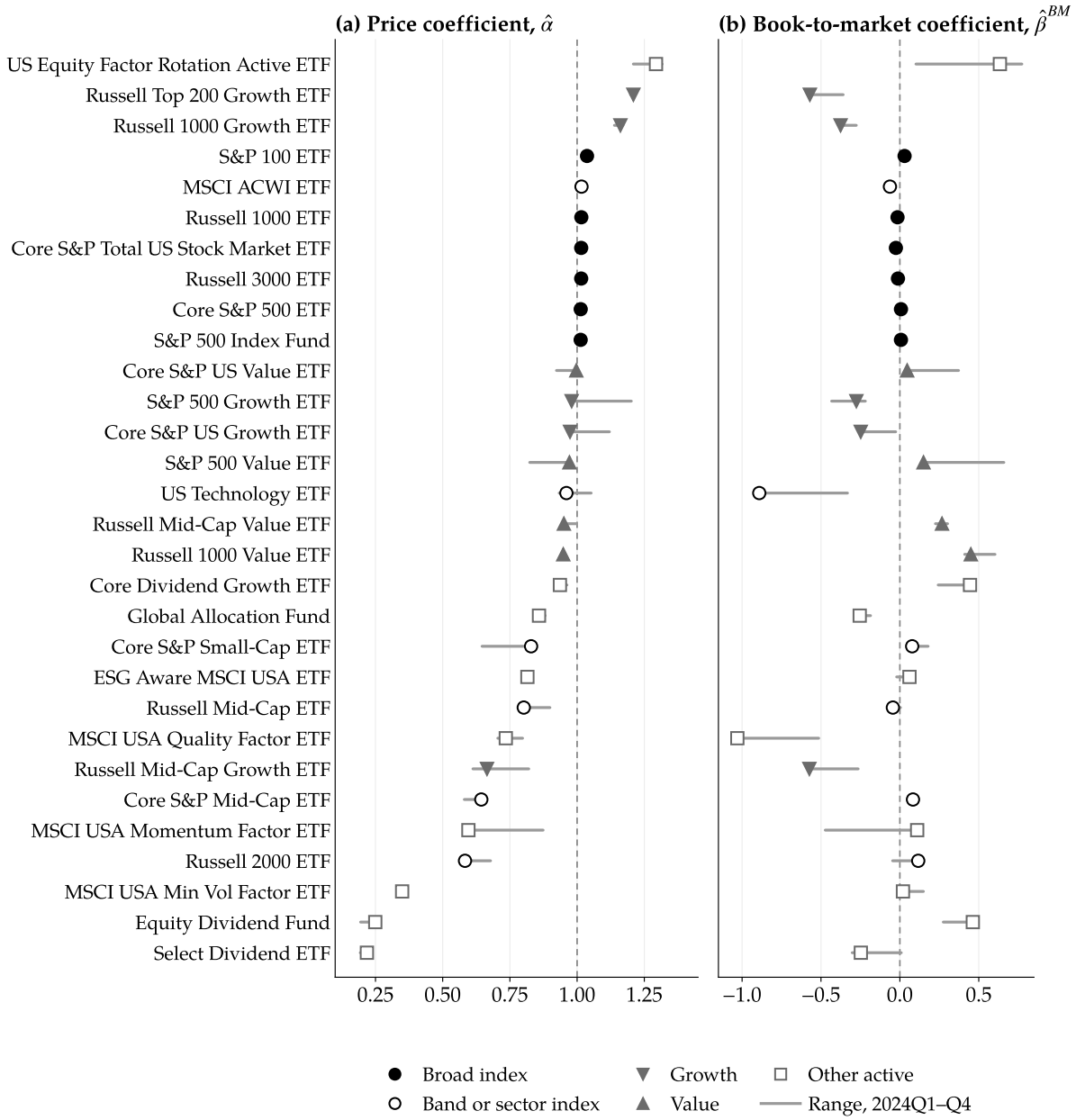

Figure 3: **Fund-level estimates for the 30 largest BlackRock and iShares funds, 2024.** Funds are the 30 largest BlackRock and iShares funds by 2024Q4 assets in the estimation sample, ordered by $\hat{\alpha}$. Markers are 2024Q4 estimates of the price coefficient $\hat{\alpha}$ (panel a) and the book-to-market coefficient $\hat{\beta}^{BM}$ (panel b). Lines are the range of each fund's estimates over the latest four quarters, 2024Q1–2024Q4. Coefficients are not winsorized. The five groups are labels for interpretation and are not used in estimation. Broad and band or sector index funds are the passive funds of footnote 8, split by index type. Growth and value funds are classified by point-in-time Lipper class, CRSP index flag and fund name (Appendix C.1). Other active funds are the remaining funds, and include all funds with "dividend" or "ESG" in their name. The dashed lines mark $\alpha = 1$ and $\beta^{BM} = 0$.

reaches 1.05 in one quarter. Under the conditional elasticity in (16), these coefficients imply that the funds increase their quantities demanded as prices rise for sufficiently small portfolio positions. Such behavior is compatible with investment strategies that favor stocks whose valuations or recent performance have risen.

These fund-level responses need not imply upward-sloping demand for the stock across its holders. A stock can be held by growth, value, index, and other funds with different responses to price. The economically relevant stock-level elasticity combines those responses using ownership weights. Section 3.3 examines whether the resulting demand is downward-sloping.

Taken together, the estimates distinguish broad index, growth, value, and other allocation strategies within the same fund family. Section 3.2.2 combines the estimated responses across each stock's holders to examine demand at the stock level.

3.2.2 Stock-level estimates

A stock's demand elasticity combines the responses of its mutual-fund holders, weighted by their ownership shares, as in (16). The resulting estimate reflects both which funds hold the stock and how those funds respond to price. I first illustrate these contributions for Tesla and Berkshire Hathaway, then examine differences across stocks and changes over time. Throughout, passive funds are conventional index funds tracking broad-market, size-band or sector indexes, and all other funds, including style, dividend and factor index funds, are non-passive.[8]

From fund to stock elasticities. Figure 4 decomposes each stock's elasticity into contributions from passive, growth, value, and other active funds. Each contribution equals the group's share of the stock's mutual-fund ownership multiplied by its ownership-weighted elasticity. The lower panels report the corresponding ownership shares.

Tesla and Berkshire reach similar estimated elasticities by 2024Q4—0.06 and 0.05, respectively—despite having different investor clienteles. Passive funds account for 63% of Tesla's mutual-fund ownership and 67% of Berkshire's, but contribute little to either stock's elasticity. Their non-passive demand comes from different strategies: growth funds remain Tesla's principal non-passive holders, accounting for 25% of its mutual-fund ownership, while Berkshire is held mainly by other active and value funds.[9] Growth funds account for only 2% of Berkshire's mutual-fund ownership.

The time series illustrates two sources of declining elasticity. Tesla's elasticity falls from 0.53 in 2010Q2 to 0.06 in 2024Q4 as passive ownership expands and the elasticity of its growth holders declines. Growth funds initially hold 85% of its mutual-fund dollars, with an average elasticity of 0.53; their average elasticity falls to 0.08 by 2024Q4. Over the same period, passive

[8] I do not split funds into a binary passive and active, because passive has many definitions. A fund tracking a Russell or S&P value or growth index is passive in that it follows a benchmark, but it also expresses a view through which stocks enter that benchmark. The categorization here is therefore coarse. Passive funds are conventional index funds tracking broad-market, size-band or sector indexes, with no style, dividend or factor marker in the fund's name. Non-passive funds are all others, including style, dividend and factor index funds, which are 17% of non-passive dollars and grow from 7% in 2010Q2 to 26% in 2024Q4.

[9] Fidelity Contrafund holds Berkshire's class A shares with an estimated elasticity near zero. Lipper reclassifies it from large-cap growth to large-cap core in 2023Q2, which moves it from the growth group to the other active group (dotted line in Figure 4).

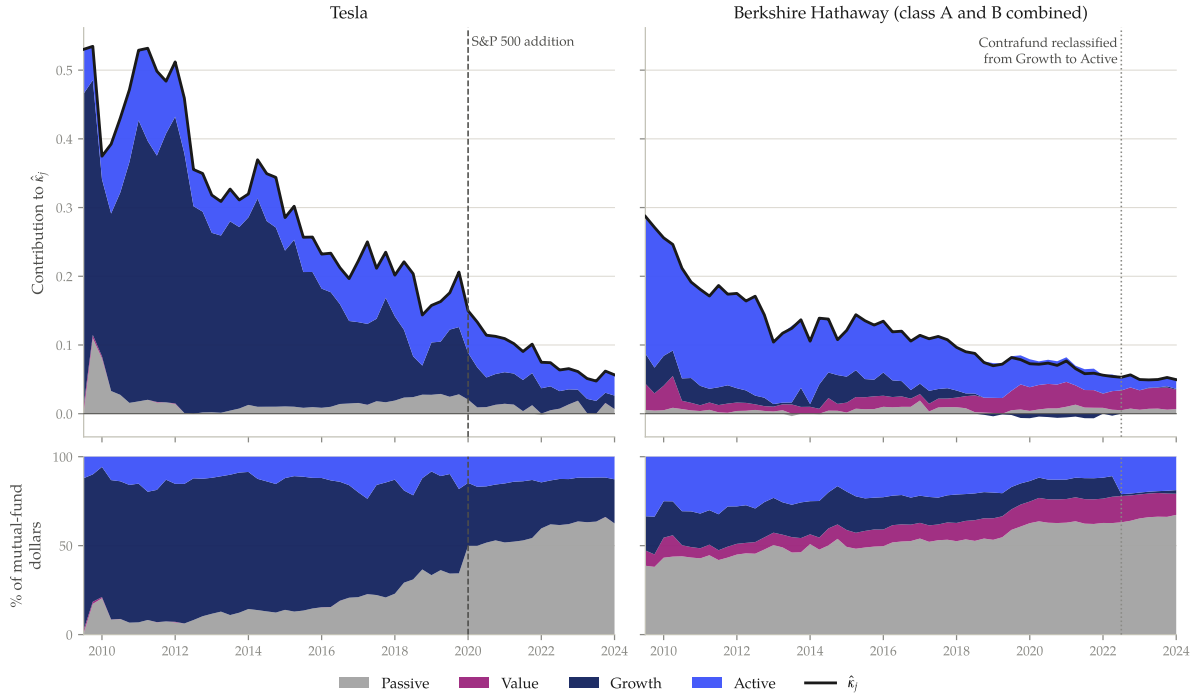

Figure 4: **Holder composition and the elasticity of Tesla and Berkshire Hathaway, 2010Q2–2024Q4.** The upper panels decompose the stock-level elasticity $\hat{\kappa}_{jt}$ (black line) into the contributions of four holder groups: each group's share of the stock's mutual-fund dollars times its dollar-weighted elasticity. The lower panels show the groups' shares of those dollars. Passive funds are defined in Footnote 8; value, growth and other active funds make up the non-passive funds. Berkshire Hathaway combines its class A and B shares. The dashed line marks Tesla's addition to the S&P 500. The dotted line marks the reclassification of Fidelity Contrafund from growth to core, which moves it from the growth to the active group.

ownership rises from 1% to 63%. Tesla's addition to the S&P 500 in December 2020 provides a particularly visible change in composition: the passive share rises from 34% to 50% that quarter, while estimated stock-level elasticity falls from 0.21 to 0.15. Berkshire's elasticity also declines, from 0.29 to 0.05. The aggregate results below examine whether these changes in composition and holder elasticities extend beyond the two examples.

The cross-section of stock elasticities. Table 1 sorts stocks by estimated elasticity each quarter, excluding stocks below the NYSE 20th size percentile. It reports time-series averages of quarterly equal-weighted means.

The least elastic stocks are predominantly large index constituents. In the lowest quintile, 78% are S&P 500 members, the average NYSE size decile is 8.4, and passive funds account for 41% of mutual-fund ownership. Index membership falls to 33% in the second quintile and 6% in the third. These patterns connect low stock-level elasticity to the large benchmark stocks held by capitalization-weighted funds.

Size, however, does not explain the full ranking. The average NYSE size decile changes little between the middle and highest elasticity quintiles, from 5.2 to 4.9. Their investor clienteles differ more substantially: growth funds account for 34% of mutual-fund ownership in the highest quintile, compared with 17% in the middle quintile. The more elastic stocks also have lower book-to-market ratios and dividend yields. These clientele and firm-characteristic differences recur across subperiods.

The ranking is not simply a consequence of passive ownership. Recomputing elasticity using only non-passive holders produces an average quarterly rank correlation of 0.93 with the original measure, and 81% of observations in each extreme quintile remain in that extreme (Table 8). Stocks with inelastic demand therefore remain relatively inelastic after conventional index funds are removed.

Nor do broad strategy labels explain these differences on their own. Growth, value, and other active funds have similar overall dollar-weighted elasticities, while the elasticities of their holdings vary substantially across stocks. For example, the average elasticity of growth holders ranges from 0.34 to 0.80 across the quintiles of non-passive elasticity. The cross-section therefore contains variation in estimated demand responses within strategy groups as well as differences in the groups holding each stock.

Low non-passive elasticity also remains associated with index membership: 65% of stocks in its lowest quintile belong to the S&P 500, compared with 1% in the highest. This pattern is consistent with benchmark-sensitive demand extending beyond conventional index funds (Pavlova and Sikorskaya, 2023), although the analysis does not directly measure funds' benchmarks.

The cross-sectional differences persist over time. Stocks sorted in the current quarter already exhibit an elasticity spread of 0.34 four quarters earlier, compared with 0.38 in the sort quarter. The high-minus-low spread also remains close to 0.38 across the three subperiods. Levels decline across all quintiles, preserving substantial differences between stocks.

Table 1: **Stocks by quintile of estimated elasticity.**

	κ_j quintile				
	Low	2	3	4	High
Elasticity					
κ_j	0.256	0.394	0.485	0.549	0.639
κ_j four quarters earlier (same stocks)	0.278	0.411	0.497	0.553	0.623
κ_j, 2010–2014	0.310	0.443	0.528	0.594	0.684
κ_j, 2015–2019	0.248	0.394	0.499	0.565	0.647
κ_j, 2020–2024	0.212	0.346	0.429	0.492	0.589
Holder base					
Mutual-fund ownership (% of shares)	28.79	31.74	36.89	41.04	41.97
Number of screened fund holders	592	345	276	276	258
Passive funds (% of mutual-fund $)	40.59	37.07	36.17	33.09	27.15
Growth funds (%)	15.25	18.11	17.02	21.39	34.41
Value funds (%)	13.08	14.56	15.03	12.92	7.90
Other active funds (%)	31.09	30.26	31.79	32.59	30.55
Firm					
S&P 500 members (%)	77.68	33.40	5.63	1.18	0.14
NYSE size decile	8.44	6.24	5.16	5.04	4.91
Market cap ($bn)	55.7	7.5	3.3	2.9	2.6
BE/ME	0.492	0.580	0.601	0.512	0.379
Momentum decile	5.82	5.70	5.80	5.91	6.20
Profitability	0.357	0.218	0.251	0.269	0.279
Investment	0.119	0.158	0.136	0.151	0.208
Market beta	1.10	1.28	1.29	1.26	1.25
Dividend yield (%)	1.84	1.49	1.41	1.16	0.80
Return volatility (annualized %)	33.80	40.80	39.55	37.74	39.02
Stock-quarters	22,790	22,757	22,752	22,757	22,778

Notes. Stocks in the stock-level panel (at least ten holders passing the fund screen), 2010Q2–2024Q4, excluding micro-caps (market capitalization below the NYSE 20th percentile). Stocks are sorted into quintiles of κ_j within each quarter. Each cell is the time-series average of quarterly equal-weighted means. Fund groups use the point-in-time Lipper classification, and fund shares are shares of the dollars held by screened mutual funds.

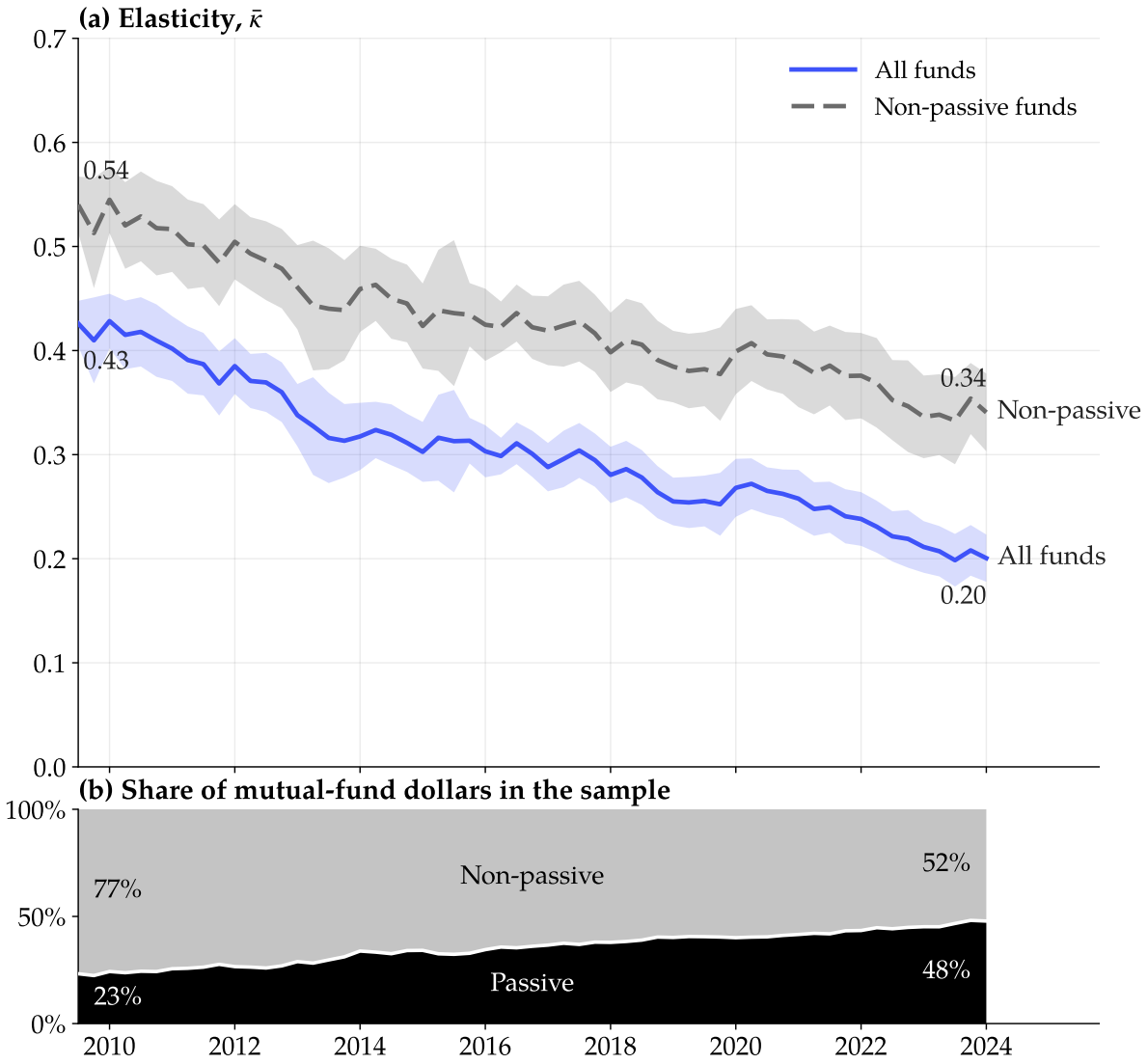

Figure 5: **Estimated demand elasticity and the passive share, 2010Q2–2024Q4.** Panel (a) plots the mutual-fund-dollar-weighted average of the stock-level elasticity κ_{jt}, for all funds and for non-passive funds. Panel (b) stacks the passive and non-passive shares of the dollars held by funds in the sample. Footnote 8 defines the two groups. Coefficients are winsorized within quarter at the 1st and 99th percentiles. The shaded areas are 95% bands for the sampling uncertainty in the fund coefficients, under the covariance assumptions of Appendix D; a few very large stocks dominate their variance.

The aggregate trend. Figure 5 shows a broad decline in the elasticity of mutual-fund demand. The mutual-fund-dollar-weighted mean falls from 0.43 in 2010Q2 to 0.20 in 2024Q4, averaging 0.30 over the sample. This decline accompanies an increase in conventional passive funds' share of sample holdings from 23% to 48%.

Changing passive ownership explains only part of the decline. Elasticity among non-passive holders also falls, from 0.54 to 0.34. Even after excluding style, dividend, and factor index funds from that group, its elasticity declines from 0.57 to 0.40. The downward trend therefore extends beyond the expansion of conventional index investing.

The decomposition attributes roughly half of the overall decline to changes in holder-group composition and half to changes in elasticities within groups. The latter includes changes in which funds hold particular stocks as well as changes in those funds' estimated responses. Together, these results show that mutual-fund demand has become less elastic through both changing ownership composition and changes within the holder groups.

3.3 The sign and stability of demand estimates

Upward-sloping demand estimates and unstable coefficients have raised doubts about the suitability of structural demand systems for financial markets. Recent critiques draw on these empirical patterns to question what the estimated demand curves identify (Fuchs et al., 2025a,b; van Binsbergen et al., 2026). This section shows that both patterns depend substantially on how discrepancies in holdings enter estimation. Some evidence interpreted as a limitation of the demand-system framework may therefore reflect the estimator used to implement it.

I re-estimate every fund-quarter using the ratio moment, retaining the same demand specification, instrument, screening rules, and winsorization procedure; neither estimator imposes a sign restriction. Section 2 showed why the ratio moment can amplify departures among small positions. Here, I examine the empirical consequences: whether the resulting estimates reproduce known index-fund allocation rules, whether demand is downward-sloping at the stock level, and whether fund-level coefficients persist over time.

Broad index funds provide a benchmark with a known allocation rule. A fund tracking a capitalization-weighted index should have a price coefficient near one and a quantity-demand elasticity near zero.[10] The level moment recovers an average elasticity of -0.001. The ratio moment instead produces an average of -0.082, with the average elasticity negative in every one of the 59 quarters. In this setting, the negative estimates attribute systematic buying in response to price increases to funds whose mandate is to track capitalization weights. The benchmark therefore identifies a direction of error: the ratio moment makes these funds' demand appear more upward-sloping than their allocation rule implies.

The same difference between moments has material consequences at the stock level (Table 2). Excluding micro-caps, negative estimated elasticity occurs in only three of approximately 114,000 stock-quarters under the level moment. Under the ratio moment, it occurs in 4.5% of stock-quarters, representing 4.0% of mutual-fund assets. By 2024Q4, the ratio estimates

[10] Size-band and sector index funds follow market capitalization only within their band or sector, so value weighting over the rolling universe is not their benchmark. Their average elasticity is 0.26 under the level moment and -0.08 under the ratio moment.

Table 2: **Level and ratio estimates compared.**

	Level	Ratio
Upward-sloping stock-quarters, % of count	0.00	4.54
Upward-sloping stock-quarters, % of mutual-fund AUM	0.01	4.00
Upward-sloping stocks in 2024Q4, % of count	0.00	16.52
Upward-sloping stocks in 2024Q4, % of mutual-fund AUM	0.00	10.04
Average broad index fund's κ	-0.001	-0.082

Notes. Fund-quarters passing the screen of Section 3.1 on each moment's own estimates, 2010Q2–2024Q4. Stocks are in the stock-level panel (at least ten holders passing the fund screen), excluding micro-caps (market capitalization below the NYSE 20th percentile). Upward-sloping means $\kappa_{jt} < 0$, with coefficients winsorized within quarter at the 1st and 99th percentiles. Broad index funds are passive funds (footnote 8) in the Lipper S&P 500, large-cap core and multi-cap core classes. Their κ is dollar-weighted within quarter and averaged across quarters.

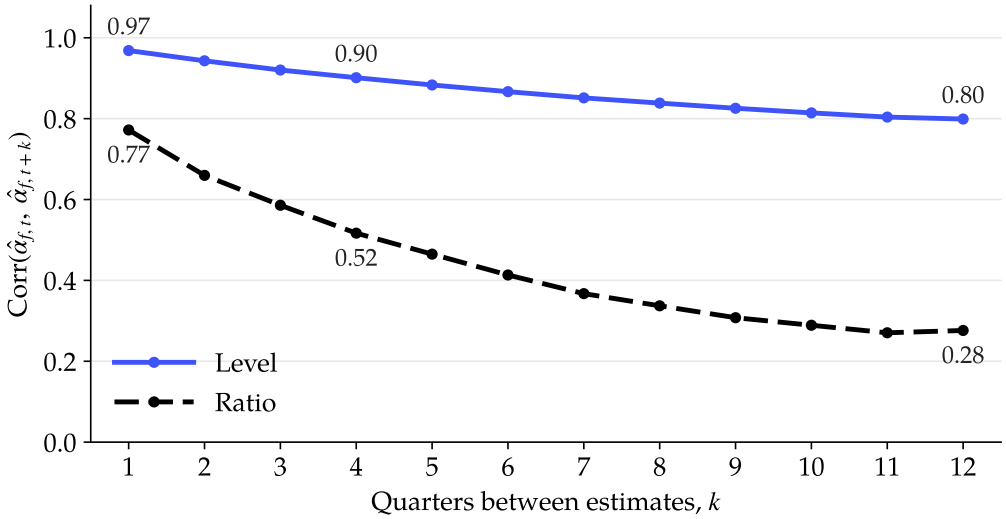

Figure 6: **Persistence of fund price coefficients.** Correlation across funds of $\hat{\alpha}$ at t and $t+k$, computed each quarter and averaged across quarters. Equal-weighted, not winsorized, same fund pairs under both moments.

imply upward-sloping demand for 16.5% of stocks and 10.0% of assets; the level estimates imply none. These differences arise without imposing downward-sloping demand on either individual funds or stocks.

The growing incidence of negative ratio estimates reflects a shift in their distribution toward zero. Between 2010 and 2024, the equal-weighted mean of stock-level elasticity falls from 0.43 to 0.11 under the ratio moment, while its cross-sectional standard deviation changes little, from 0.13 to 0.12. As the mean approaches zero, more of the distribution crosses into negative elasticity. The equal-weighted mean also declines under the level moment, from 0.53 to 0.35, but remains substantially above zero. The choice of moment therefore affects the level of estimated demand elasticity as well as the incidence of negative estimates.

The persistence of fund-level coefficients provides complementary evidence (Figure 6). The correlation of a fund's price coefficient with its value one, four, and twelve quarters later is 0.97, 0.90, and 0.80 under the level moment, compared with 0.77, 0.52, and 0.28 under the ratio moment. Unwinsorized ratio estimates also span a much wider range, from -24 to 22, compared with -1.7 to 2.5 under the level moment. Because investment mandates generally persist across quarters, the greater persistence of level estimates is consistent with less estimation noise around those allocation rules. Persistence alone cannot establish accuracy, but it supports the index-fund evidence that the moment materially affects the behavior recovered from holdings.

These results show that upward-sloping stock-level demand and substantial quarter-to-quarter instability are sensitive to the estimating moment. The index-fund benchmark demonstrates a systematic discrepancy under ratio estimation, while the broader comparison shows that changing the moment largely removes negative stock-level elasticities and produces more persistent fund coefficients. These patterns motivate addressing the treatment of holding discrepancies before interpreting them as limitations of the demand model.

4 Out-of-Sample Validation

Each estimated price coefficient is a claim about behavior: it predicts how a fund reallocates when prices change while its demand remains otherwise unchanged. Section 3 showed that the estimates are economically interpretable and stable. This section tests their behavioral predictions directly: when prices move, do funds trade as their previously estimated demand curves imply?

Comparisons with elasticities inferred from index events (Shleifer, 1986; Harris and Gurel, 1986; Petajisto, 2009), flow-induced trading (Lou, 2012), or other demand shifts (Gabaix and Koijen, 2021) need not evaluate the same object. Those designs generally allow other prices to adjust, whereas a structural own-price response holds them fixed. Similar magnitudes can therefore conceal different demand responses, as emphasized by Fuchs et al. (2025a). A demand system also makes predictions for individual investors, providing a further dimension along which it can be evaluated.

I construct an out-of-sample test that compares the trading of funds facing the same price changes. The coefficients are estimated from holdings two quarters before the trading period.

Stock-quarter fixed effects absorb components common to holders of the same stock, while fund-quarter fixed effects absorb portfolio-wide components. Under the logit's substitution structure, the latter include the effects of other price changes. The comparison therefore isolates differences in funds' predicted responses along their demand curves.

With price changes instrumented by flow-induced trading, funds implement about two-thirds of the predicted differences in trading within a quarter. The predicted ordering is evident across the distribution of funds, while departures from full adjustment are concentrated among funds with the most extreme estimated coefficients.

Section 4.1 sets out the test and what identifies it, Section 4.2 reports the results, and Section 4.3 asks whether the estimated coefficients predict trading only because they sort funds on how active they are.

4.1 Design and identification

The prediction. Consider fund f between the quarter-ends $t-1$ and t, and let Δp_{jt} be the change in stock j's log market equity. If the fund did not trade, its weights would drift with prices. To first order,

$$\log w^{d}_{fjt} - \log w_{fj,t-1} = \Delta p_{jt} - \Delta \bar{p}_{ft}, \qquad \Delta \bar{p}_{ft} = \sum_{k} w_{fk,t-1}\, \Delta p_{kt}, \tag{17}$$

where w^{d}_{fjt} is the drift weight and $\Delta \bar{p}_{ft}$ the return on the fund's portfolio.[11] Under the demand system of Section 3.1, the same price changes move the fund's desired weights, holding its coefficients, its latent demand and the stocks' other characteristics fixed, by

$$\Delta \log w^{*}_{fjt} = \alpha_f \big(\Delta p_{jt} - \Delta \bar{p}_{ft}\big). \tag{18}$$

The predicted trade is the difference between the two:

$$\text{Trade}^{*}_{fjt} \equiv \Delta \log w^{*}_{fjt} - \big(\log w^{d}_{fjt} - \log w_{fj,t-1}\big) = (\alpha_f - 1)\big(\Delta p_{jt} - \Delta \bar{p}_{ft}\big). \tag{19}$$

A value-weighted index fund, with $\alpha_f = 1$, does not trade. A fund with $\alpha_f < 1$ sells stocks that outperformed its portfolio and buys those that underperformed, and the more so the lower its α_f.

Equation (19) holds the stock's other characteristics fixed. Price changes also affect book-to-market and, when they enter its measurement window, momentum: with book equity fixed, a price increase lowers log book-to-market one for one. A fund's tastes for these characteristics therefore also generate trading when prices move. These responses are part of the fund's demand but not the price slope that α_f measures. The controls in Table 3 therefore include, for each of the two characteristics, its change times the fund's estimated coefficient on it, instrumented separately (Appendix C.4 for momentum), so that γ measures the response along the

[11] Weights drift with the split-adjusted share price, whose log change is $\Delta p_{jt} - \Delta s_{jt}$, where Δs_{jt} is the change in log shares outstanding; desired weights respond to market equity, the characteristic in the demand system. The exact predicted trade is therefore $\alpha_f(\Delta p_{jt} - \Delta \bar{p}_{ft}) - \big[(\Delta p_{jt} - \Delta s_{jt}) - (\Delta \bar{p}_{ft} - \Delta \bar{s}_{ft})\big] = (\alpha_f - 1)(\Delta p_{jt} - \Delta \bar{p}_{ft}) + (\Delta s_{jt} - \Delta \bar{s}_{ft})$, with $\Delta \bar{s}_{ft}$ weighted like $\Delta \bar{p}_{ft}$. The second term does not involve α_f: Δs_{jt} is common to all holders of the stock and $\Delta \bar{s}_{ft}$ to the fund's portfolio, so the stock-quarter and fund-quarter effects of (20) absorb it.

price coefficient alone.[12]

The estimating equation. Let $\text{Trade}_{fjt} = w_{fjt}/w^d_{fjt} - 1$ be the fund's trade in stock j relative to drift. I estimate

$$\text{Trade}_{fjt} = \gamma \widehat{\text{Trade}}_{fjt} + \sum_{\ell} \zeta_\ell \, \mathbf{1}[L(f) = \ell] \, \Delta p_{jt} + x'_{fjt}\delta + \phi_{ft} + \psi_{jt} + e_{fjt}, \tag{20}$$

where $\widehat{\text{Trade}}_{fjt}$ uses the fund's estimate $\hat{\alpha}_f$ from quarter $t-2$, shrunk toward its own estimates of the three preceding quarters, so that no information after $t-2$ enters it. $L(f)$ is the fund's Lipper class, so each class has its own response to price and γ measures whether a fund's own estimate predicts its trading beyond the average of its style. x_{fjt} collects the controls of Table 3, and ϕ_{ft} and ψ_{jt} are fund-quarter and stock-quarter fixed effects. The sample is the positions held at both $t-1$ and t, weighted by their drift weights, over the trading quarters 2012Q3–2024Q4. The market-equity component of the model's trading prediction has a coefficient of one under full adjustment, $\gamma = 1$, as do the momentum and book-to-market terms among the controls.

What the comparison holds fixed. The predicted trade has two parts,

$$\text{Trade}^*_{fjt} = (\alpha_f - 1)\,\Delta p_{jt} - (\alpha_f - 1)\,\Delta \bar{p}_{ft}.$$

The stock-quarter effect ψ_{jt} absorbs components of trading common to holders of stock j, including common responses to firm news. The fund-quarter effect ϕ_{ft} absorbs portfolio-wide components, including the term $-(\alpha_f - 1)\Delta\bar{p}_{ft}$. Under the logit, other stocks' price changes affect demand for each holding through this common portfolio term.[13] Their effects are therefore absorbed even when the observed price changes reflect general-equilibrium adjustments.

Identification comes from differences across funds holding the same stock in the same quarter: do funds with lower estimated price coefficients trade more against its price change, as their demand curves predict? The test evaluates these investor-level quantity responses using previously estimated coefficients. This differs from inferring an elasticity from the equilibrium price impact of a demand shift.

The separation of own-price and cross-price responses relies on the logit's independence of irrelevant alternatives. With richer substitution patterns, a peer stock's price could affect demand for j differently from demand for the fund's other holdings. Such effects would remain in the residual.

Why the price change needs an instrument. The fixed effects remove everything common to a stock or to a fund. They do not remove a change in demand that is specific to the fund and correlated with the stock's return. If funds with a low $\hat{\alpha}_f$ also tend to chase recent winners, or

[12] The book-to-market term matters little for γ. It is nearly uncorrelated with the predicted trade (0.03 within the fixed effects), and without it γ is 0.671 (0.052) rather than 0.682 (0.052). Funds carry out about a quarter of the book-to-market trade, 0.259 (0.057).

[13] The own-price elasticity $\kappa_{fjt} = 1 - \alpha_f(1 - w_{fjt})$ carries the factor $1 - w_{fjt}$ because stock j's own return is part of $\Delta\bar{p}_{ft}$. The fixed effect absorbs that part too, which is immaterial except for a fund's largest positions.

receive news that moves both their own trades and the price, the residual trade is correlated with Trade^*_{fjt} in a way that differs across funds, and γ is biased. Estimating $\hat{\alpha}_f$ before the trading period prevents the prediction from using contemporaneous holdings, but does not eliminate correlation between price changes and subsequent demand shocks.

Flow-induced trading as the instrument. I instrument the stock's relative return with the trading that other funds' flows force on it (Lou, 2012). Flow-induced trading in stock j is

$$\text{FIT}_{j,-S(f),t} = \sum_{g \notin S(f)} \frac{\text{shares of } j \text{ held by } g \text{ at } t-1}{\text{shares outstanding of } j \text{ at } t-1} \text{F}_{gt}, \tag{21}$$

where F_{gt} is fund g's net flow in quarter t as a fraction of its total net assets at the end of quarter $t-1$, and $S(f)$ is the set of funds that share either fund f's fund company or its Lipper class. The instrument for $\widehat{\text{Trade}}_{fjt}$ is $(\hat{\alpha}_f - 1)\left(\text{FIT}_{j,-S(f),t} - \overline{\text{FIT}}_{ft}\right)$, built in the same way as the prediction, where $\overline{\text{FIT}}_{ft}$ is the average of $\text{FIT}_{j,-S(f),t}$ over fund f's holdings, weighted like $\Delta\bar{p}_{ft}$. I use two versions of the flow: the raw flow, and the unexpected flow, which is the raw flow minus its forecast from the fund's past flows and past performance, following Lou (2012). The unexpected flow is the main version.

The exclusion restriction is that flow-induced price pressure is unrelated to the focal fund's stock-specific demand shock, conditional on the controls and fixed effects. Two features support this restriction. Unexpected flows remove the component predicted by past flows and performance. Excluding the fund's own company and Lipper class reduces direct exposure to common company-level and style-level demand shocks.

4.2 Results

Table 3 reports the estimates. The columns add the controls one at a time: the fixed effects only, then Lipper class $\times$ price change, own-company FIT, and the stock characteristics. The sample has 20.9 million positions of 5,974 funds in 3,413 stocks over 50 quarters, the same in every column.

Funds trade in the direction their estimated demand curves predict, in every column. By OLS (panel A), funds carry out about a quarter of the predicted difference in trading, from 0.235 with the fixed effects only to 0.300 with all controls. Instrumenting the price change raises the estimate. With unexpected-flow FIT and all controls, panel C, column (4), γ is 0.682 with a standard error of 0.052, against 0.300 by OLS: funds carry out about two-thirds of the predicted difference within the quarter. The first-stage F statistics range from 63 to 91, so the instrument is strong in every column.

The stock characteristics matter for the unexpected-flow estimate, which rises from 0.429 to 0.682 when they are added (panel C, columns (3) and (4)). Most of the rise comes from the momentum characteristic. Funds' estimated price coefficients and their estimated tastes for momentum are negatively correlated, so without that control part of a fund's response to recent returns is attributed to its price coefficient (Appendix C.4). The raw-flow estimates in panel B are higher, 0.606 and 0.776 in columns (3) and (4). Raw flows partly follow past returns, which the unexpected flow removes, so I take the unexpected-flow estimate as the main one.

Table 3: **Trading response to the demand system's predicted trade.**

	(1)	(2)	(3)	(4)
Panel A: OLS				
Predicted trade, γ	0.235	0.264	0.264	0.300
	(0.014)	(0.014)	(0.014)	(0.015)
Panel B: IV, raw-flow FIT				
Predicted trade, γ	0.496	0.594	0.606	0.776
	(0.057)	(0.067)	(0.060)	(0.072)
First stage	3.48	2.82	3.00	3.35
	(0.44)	(0.35)	(0.38)	(0.37)
First-stage F	62.8	63.7	64.1	82.9
Panel C: IV, unexpected-flow FIT				
Predicted trade, γ	0.378	0.451	0.429	0.682
	(0.046)	(0.056)	(0.046)	(0.052)
First stage	4.30	3.48	3.69	4.32
	(0.53)	(0.43)	(0.45)	(0.45)
First-stage F	66.1	64.4	66.2	90.9
Fund-quarter, stock-quarter FE	Yes	Yes	Yes	Yes
Lipper class × price change	No	Yes	Yes	Yes
Own-company FIT	No	No	Yes	Yes
Stock characteristics	No	No	No	Yes
Observations	20,941,098	20,941,098	20,941,098	20,941,098
Funds	5,974	5,974	5,974	5,974
Stocks	3,413	3,413	3,413	3,413
Quarters	50	50	50	50

Notes. Each column estimates $\text{Trade}_{fjt} = \gamma \widehat{\text{Trade}}_{fjt} + \text{controls} + \phi_{ft} + \psi_{jt} + e_{fjt}$, where $\text{Trade}_{fjt} = w_{fjt}/w^d_{fjt} - 1$ is fund f's trade in stock j relative to its drift weight and $\widehat{\text{Trade}}_{fjt} = (\hat{\alpha}_f - 1)(\Delta p_{jt} - \Delta \bar{p}_{ft})$ is the demand system's predicted trade, with $\hat{\alpha}_f$ estimated at $t-2$ and shrunk toward the fund's own estimates of the three preceding quarters. Under full adjustment $\gamma = 1$. The sample is the positions held at both $t-1$ and t by funds that pass the estimation screen at $t-2$ and have a shrinkage baseline, in stocks above the NYSE 20th size percentile at $t-1$, for trading quarters 2012Q3–2024Q4; it is the same in every column. Each position is weighted by its drift weight w^d_{fjt}: the fund's shares of stock j at $t-1$ valued at t prices, as a share of the fund's full $t-1$ portfolio valued at t prices. A fund-quarter's positions therefore weigh in total the share of its portfolio it continues to hold, and no observation is dropped. Panels B and C instrument $\widehat{\text{Trade}}_{fjt}$ with $(\hat{\alpha}_f - 1)$ times the flow-induced trading (FIT) in stock j of funds outside fund f's company and Lipper class, built from raw flows (Panel B) or from unexpected flows, flows minus the forecast of Lou (2012) (Panel C). Lipper class × price change gives each class its own response to price. Own-company FIT, built from the same flows as the instrument, controls for the part the instrument leaves out. Stock characteristics are size, book-to-market, momentum, return volatility, illiquidity, mutual-fund ownership and the index-fund share of mutual-fund holdings, each interacted with $(\hat{\alpha}_f - 1)$, and two terms for the characteristics that move with price: the change in the stock's 12–2 month momentum times the fund's estimated taste for momentum, instrumented with unexpected-flow FIT over the same months, and the change in the stock's log book-to-market times the fund's estimated book-to-market coefficient, instrumented with that coefficient times the panel's FIT, with the corresponding own-company FIT as a control. Their coefficients, which also equal one under full adjustment, are not reported. No past-quarter return is included. Standard errors (in parentheses) are clustered by fund, stock and quarter. The first-stage F is the effective F for the price instrument.

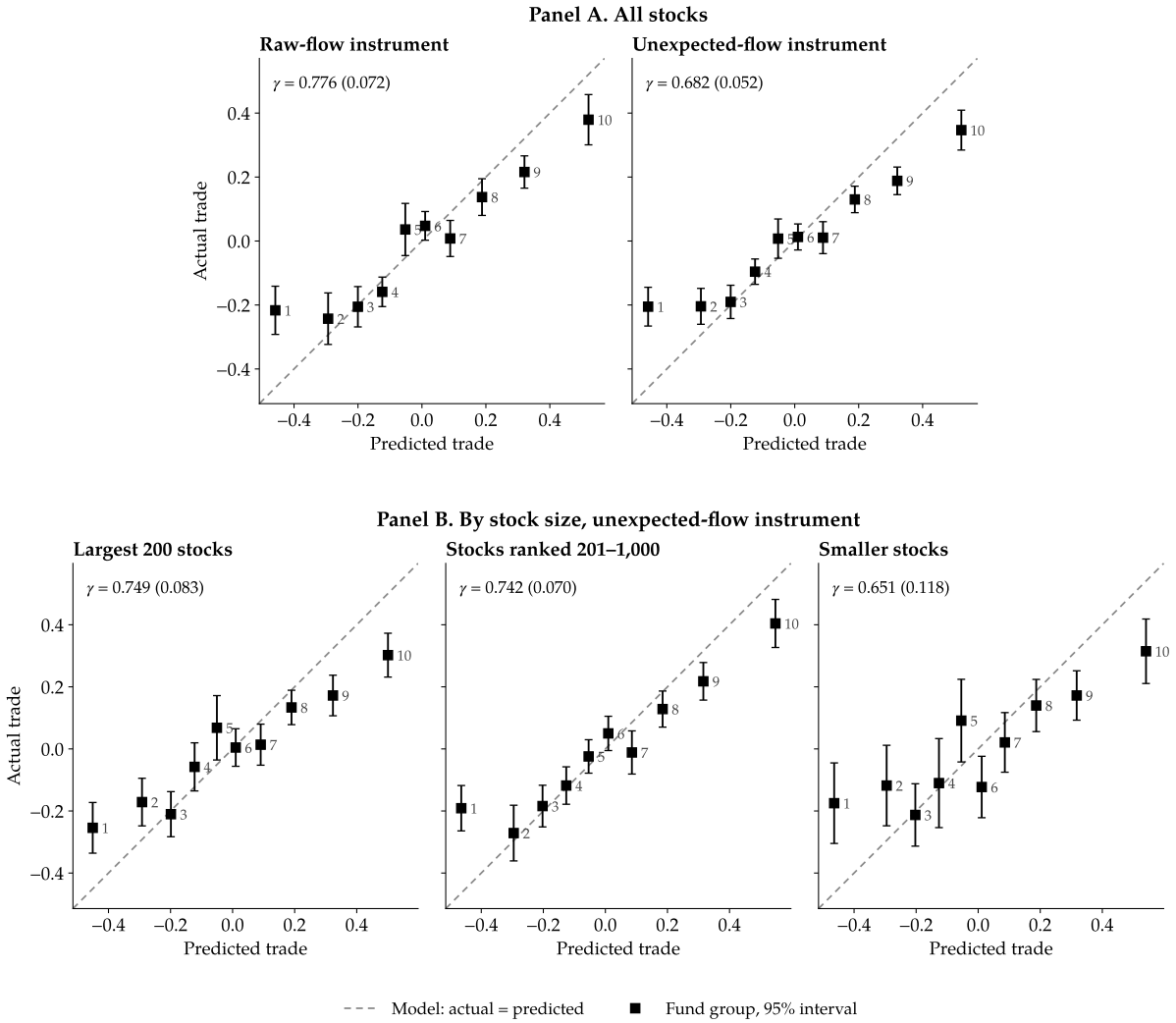

Figure 7: **Predicted and actual trades across fund groups.** Each quarter, funds are ranked on $\hat{\alpha}_f - 1$ net of their Lipper-class mean and split into ten groups. The horizontal axis is the group's mean predicted trade, weighted by the share of each fund's portfolio that continues to be held. The vertical axis is the group's actual trading response to an instrumented price shock, estimated group by group in the specification of Table 3 with all controls. Both axes are relative to the ten-group mean, because the stock-quarter fixed effects leave only differences across funds. Bars are 95% intervals. The dashed line is the model's prediction, actual equal to predicted. The number in each graph is γ, the coefficient on the predicted trade. Panel A uses all stocks: the left graph uses the raw-flow instrument (Table 3, panel B, column (4)) and the right graph the unexpected-flow instrument (panel C, column (4)). Panel B repeats the unexpected-flow specification separately for the largest 200 stocks, stocks ranked 201–1,000, and smaller stocks, ranked by market capitalization at $t-1$, each with its own fixed effects (Table 9, panel E). Standard errors are clustered by fund, stock and quarter.

Figure 7 shows where the shortfall from one comes from. Funds are sorted each quarter into ten groups by their estimated coefficient, and each group's actual trade is plotted against its predicted trade. The groups line up in the predicted order in both graphs of Panel A. The middle groups lie close to the 45-degree line, and the shortfall comes from the tails: the funds with the lowest and the highest estimated coefficients respond less than predicted.

The estimate is stable across the choices in its construction (Appendix E, Table 9). Seven alternative constructions of the unexpected flow give estimates from 0.61 to 0.69, and changing the funds whose trading is left out of the instrument gives 0.66 to 0.67. Using the unshrunk estimates of the fund's tastes gives 0.61. By subperiod, the estimate is 0.88 in 2012Q3–2016Q4 and 0.62 in 2017–2024, and excluding 2020Q2–2021Q1 gives 0.72.

Panel B of Figure 7 repeats the unexpected-flow calibration separately for the largest 200 stocks, the stocks ranked 201 to 1,000, and smaller stocks, each with its own fixed effects. In each group of stocks, actual trades rise with predicted trades, and the funds with the most extreme estimates again respond less than predicted. The coefficient is 0.75 for the largest 200 stocks, 0.74 for the next 800 and 0.65 for smaller stocks, where the group estimates are least precise. None is far from the pooled estimate of 0.68: funds act on their estimated demand curves in small stocks as well as large ones.

The relative test therefore supports the cross-fund differences in the estimated demand curves: funds with lower estimated coefficients trade more against the same price change, and they carry out about two-thirds of the predicted difference within a quarter. The test does not speak to the level of the coefficients, which the stock-quarter effects remove.

4.3 Absorption test

The predicted trade has the form of a rebalancing rule: a fund with $\hat{\alpha}_f < 1$ sells stocks that outperformed its portfolio and buys those that underperformed. Any fund that actively trades against the drift in its weights does the same. The estimated coefficient could therefore predict trading only because it sorts funds on how active they are, rather than because it measures the slope of their demand curves. I test this by adding fund traits that measure activeness and scale to the specification of Table 3, panel C, column (4), and asking whether they absorb the predictive content of $\hat{\alpha}_f$.

The traits are active share (Cremers and Petajisto, 2009); factor-residual volatility, the annualized standard deviation of the fund's monthly four-factor residual over the preceding twelve months; and fund size, the log of its assets under management. All are measured at $t-2$, when $\hat{\alpha}_f$ is estimated. Each trait enters as its standardized value within the quarter times the relative price change, instrumented in the same way as the predicted trade, so its coefficient δ is the difference in the trading response per standard deviation of the trait. Table 4 reports, for each trait, δ without and with $\hat{\alpha}_f$, and γ with the trait, for all funds in panel A and for funds not labelled as index funds in panel B.

Active share is the only trait that moves γ. On its own it predicts the response: a fund with active share one standard deviation higher trades more against relative price changes, with $\delta = -0.198$ (0.020) in panel A. With $\hat{\alpha}_f$ included, active share keeps about half of that, -0.110 (0.020), and γ falls from 0.69 to 0.49. Factor-residual volatility and fund size predict the

Table 4: **The price coefficient against fund activeness and size.**

	Trait coefficient δ		Coefficient on $\hat{\alpha}_f - 1$, γ		
	Without $\hat{\alpha}_f$	With $\hat{\alpha}_f$	With the trait	Change	Conditional
	(1)	(2)	(3)	(4)	F
Panel A: All funds					
$\hat{\alpha}_f - 1$ only			0.691		96.0
			(0.053)		
Active share	−0.198	−0.110	0.486	−0.205	82.4
	(0.020)	(0.020)	(0.056)	(0.035)	
Factor-residual volatility	−0.067	−0.018	0.678	−0.013	95.6
	(0.013)	(0.015)	(0.056)	(0.012)	
Fund size	0.035	−0.005	0.688	−0.002	93.8
	(0.011)	(0.012)	(0.054)	(0.008)	
All three traits			0.490	−0.200	82.0
			(0.057)	(0.035)	
20.0 million positions, 5,793 funds, 130,151 fund-quarters					
Panel B: Non-index funds					
$\hat{\alpha}_f - 1$ only			0.616		95.2
			(0.067)		
Active share	−0.156	−0.079	0.464	−0.152	87.3
	(0.020)	(0.018)	(0.073)	(0.033)	
Factor-residual volatility	−0.040	−0.005	0.613	−0.002	96.3
	(0.015)	(0.016)	(0.070)	(0.013)	
Fund size	0.009	−0.016	0.620	0.005	93.7
	(0.012)	(0.013)	(0.068)	(0.008)	
All three traits			0.469	−0.146	86.7
			(0.073)	(0.033)	
10.7 million positions, 4,610 funds, 100,752 fund-quarters					

Notes. Every regression is the specification of Table 3, Panel C, column (4): Trade_{fjt} on fund traits times the relative price change $\Delta p_{jt} - \Delta \bar{p}_{ft}$, with the momentum and book-to-market terms, stock characteristics $\times(\hat{\alpha}_f - 1)$, Lipper class $\times$ price change, own-company FIT, fund-quarter and stock-quarter fixed effects, and drift weights. The demand system's trait enters as $(\hat{\alpha}_f - 1)(\Delta p_{jt} - \Delta \bar{p}_{ft})$, its predicted trade $\widehat{\text{Trade}}_{fjt}$, so $\gamma = 1$ under the model; it is instrumented with $(\hat{\alpha}_f - 1)$ times unexpected-flow FIT. Each other trait (active share, factor-residual volatility, log fund size, all at $t-2$) enters as $q_{ft}(\Delta p_{jt} - \Delta \bar{p}_{ft})$, where q_{ft} is the trait standardised within the quarter over the panel's funds (weighted by continuing share), instrumented with q_{ft} times the same FIT, with q_{ft} times own-company FIT as a control; δ is the response per standard deviation of the trait. Each row adds one trait: column (1) is its δ with every $\hat{\alpha}_f$ term removed (the predicted trade, the characteristics interactions and own-company FIT $\times(\hat{\alpha}_f - 1)$); columns (2) and (3) are δ and γ when the trait and $\hat{\alpha}_f$ enter together; column (4) is the change in γ from the first row of the panel. The last row of each panel adds all three traits. For comparison with δ, γ per standard deviation of $\hat{\alpha}_f - 1$ is 0.242 in panel A and 0.193 in panel B. Within quarter, $\hat{\alpha}_f - 1$ has correlation −0.64 with active share, −0.22 with factor-residual volatility and 0.25 with fund size (panel A). Panel A: all funds; panel B: funds not labelled as index funds at $t-2$. Both are restricted to fund-quarters with all three traits, which is why the first row of panel A differs slightly from Table 3, Panel C, column (4). The conditional F is the Sanderson–Windmeijer F of the $\hat{\alpha}_f$ term. Standard errors (in parentheses) are clustered by fund, stock and quarter. Trading quarters 2012Q3–2024Q4.

response on their own, -0.067 and 0.035 per standard deviation, but not once $\hat{\alpha}_f$ is included, and neither changes γ. With all three traits, γ is 0.490 (0.057). Among non-index funds, γ falls from 0.62 to 0.47 with all three traits, and active share keeps -0.079 (0.018) of its -0.156.

The estimated coefficient therefore absorbs the predictive content of factor-residual volatility and fund size, and about half of that of active share, while active share takes about 30 percent of γ in panel A. The two are closely related: within a quarter, $\hat{\alpha}_f - 1$ and active share have a correlation of -0.64, so the regression separates two overlapping fund-level gradients, and estimation error in either can shift weight between them. With all three traits controlled for, the coefficient on the predicted trade remains about one-half and is precisely estimated. The demand system's coefficients carry information about trading beyond these measures of activeness and size.

5 Conclusion

This paper shows that how portfolio holdings enter estimation materially affects the demand behavior inferred from them. Institutional portfolios contain a long tail of small positions whose implementation and reporting departures can be large relative to their fitted weights. Conventional relative-error normalization amplifies these discrepancies. Measuring them in absolute units produces economically interpretable and stable estimates without imposing sign restrictions. Estimated mutual-fund demand is substantially more elastic than conventional estimation implies, although it remains inelastic. These estimates also predict subsequent trading. In a structural test designed to separate investors' demand responses from general-equilibrium price adjustments, funds implement about two-thirds of the predicted differences in trading within a quarter. A simple demand system estimated from portfolio allocations therefore captures economically meaningful differences in how investors respond to prices.

The contribution of this paper is primarily empirical, and it is not intended to settle the broader debate. Recent work emphasizes the role of dynamics and cross-asset substitution in identifying and interpreting asset demand (e.g., He et al., 2026; Haddad et al., 2026), and incorporating these insights into estimation is a natural extension. Equally, the performance of the simple specification deserves explanation precisely because its predictions inherit the restrictive substitution structure of the logit model. One avenue is to examine whether this structure provides a useful approximation to portfolio choice when investors face benchmark incentives. Understanding the economic sources of this predictive success could help guide the development of richer demand systems for financial markets.

Taken together, these findings support a constructive role for demand systems in finance. Asset pricing has learned much from studying prices and return moments. Quantity moments drawn from portfolio holdings and trading offer a complementary source of evidence about investor behavior. The results here show that demand systems can organize this evidence into economically interpretable and testable predictions. The broader ambition is a framework disciplined jointly by prices and quantities, linking who holds assets and how they trade to the prices and returns observed in financial markets.

References

Amiti, Mary and David E. Weinstein, "How Much Do Idiosyncratic Bank Shocks Affect Investment? Evidence from Matched Bank-Firm Loan Data," *Journal of Political Economy*, 2018, *126* (2), 525–587.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

BlackRock, "iShares Total U.S. Stock Market Index Fund: Summary Prospectus," BlackRock Funds, Investor A and Institutional Shares November 2025. Dated November 28, 2025, p. 3.

Bretscher, Lorenzo, Lukas Schmid, Ishita Sen, and Varun Sharma, "Institutional Corporate Bond Pricing," *Review of Financial Studies*, 2026, *39* (3), 605–660.

Brogaard, Jonathan, Davidson Heath, and Da Huang, "ETF Sampling and Index Arbitrage," *Journal of Financial and Quantitative Analysis*, 2026, *61* (2), 547–579.

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, "Regression Discontinuity and the Price Effects of Stock Market Indexing," *Review of Financial Studies*, 2015, *28* (1), 212–246.

Chaudhary, Manav, Zhiyu Fu, and Jian Li, "Corporate Bond Multipliers: Substitutes Matter," 2026. Financial Markets Group Discussion Paper 955, June 2026; SSRN 4292266.

Chodorow-Reich, Gabriel, "The Employment Effects of Credit Market Disruptions: Firm-Level Evidence from the 2008–9 Financial Crisis," *Quarterly Journal of Economics*, 2014, *129* (1), 1–59.

Cohn, Jonathan B., Zack Liu, and Malcolm I. Wardlaw, "Count (and Count-Like) Data in Finance," *Journal of Financial Economics*, 2022, *146* (2), 529–551.

Cremers, K. J. Martijn and Antti Petajisto, "How Active Is Your Fund Manager? A New Measure That Predicts Performance," *Review of Financial Studies*, 2009, *22* (9), 3329–3365.

Davis, Carter, Mahyar Kargar, and Jiacui Li, "Why Do Portfolio Choice Models Predict Inelastic Demand?," *Journal of Financial Economics*, 2025, *172*, 104096.

Erickson, Timothy and Toni M. Whited, "Measurement Error and the Relationship between Investment and q," *Journal of Political Economy*, 2000, *108* (5), 1027–1057.

Fama, Eugene F. and Kenneth R. French, "Dissecting Anomalies," *Journal of Finance*, 2008, *63* (4), 1653–1678.

Fuchs, William, Satoshi Fukuda, and Daniel Neuhann, "Demand-System Asset Pricing: Theoretical Foundations," 2025. Working paper, April 16, 2025; SSRN 4672473.

_ , _ , and _ , "A Trilemma for Asset Demand Estimation," 2025. Working paper, SSRN 5355589; version of May 8, 2026.

Gabaix, Xavier and Ralph S. J. Koijen, "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," NBER Working Paper 28967, National Bureau of Economic Research 2021.

Gârleanu, Nicolae and Lasse Heje Pedersen, "Dynamic Trading with Predictable Returns and Transaction Costs," *Journal of Finance*, 2013, *68* (6), 2309–2340.

Gonçalves, Andrei S., Chen Xue, and Lu Zhang, "Aggregation, Capital Heterogeneity, and the Investment CAPM," *Review of Financial Studies*, 2020, *33* (6), 2728–2771.

Graves, Daniel, "What Lies Beneath Zero: Censoring, Demand Estimation, and Hidden Beliefs," 2024. Working paper (job market paper), Yale University, August 2024.

Guimarães, Paulo, Octávio Figueiredo, and Douglas Woodward, "A Tractable Approach to the Firm Location Decision Problem," *Review of Economics and Statistics*, 2003, *85* (1), 201–204.

Haddad, Valentin, Paul Huebner, and Erik Loualiche, "How Competitive Is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing," *American Economic Review*, 2025, *115* (3), 975–1018.

_ , Zhiguo He, Paul Huebner, Péter Kondor, and Erik Loualiche, "Causal Inference for Asset Pricing," NBER Working Paper 35413, National Bureau of Economic Research July 2026.

Harris, Lawrence and Eitan Gurel, "Price and Volume Effects Associated with Changes in the S&P 500 List: New Evidence for the Existence of Price Pressures," *Journal of Finance*, 1986, *41* (4), 815–829.

He, Zhiguo, Péter Kondor, and Jessica S. Li, "Demand Elasticity in Dynamic Asset Pricing," 2026. NBER Working Paper 34450, revised July 2026; first version 2025.

Jansen, Kristy A. E., Wenhao Li, and Lukas Schmid, "Granular Treasury Demand with Arbitrageurs," CEPR Discussion Paper 21079, Centre for Economic Policy Research 2026. Earlier version: NBER Working Paper 33243, December 2024.

Kaul, Aditya, Vikas Mehrotra, and Randall Morck, "Demand Curves for Stocks Do Slope Down: New Evidence from an Index Weights Adjustment," *Journal of Finance*, 2000, *55* (2), 893–912.

Khwaja, Asim Ijaz and Atif Mian, "Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market," *American Economic Review*, 2008, *98* (4), 1413–1442.

Koijen, Ralph S. J. and Motohiro Yogo, "A Demand System Approach to Asset Pricing," *Journal of Political Economy*, 2019, *127* (4), 1475–1515.

_ and _ , "Exchange Rates and Asset Prices in a Global Demand System," *Journal of Political Economy*, 2026, *134* (8), 2441–2487.

_ , François Koulischer, Benoît Nguyen, and Motohiro Yogo, "Inspecting the Mechanism of Quantitative Easing in the Euro Area," *Journal of Financial Economics*, 2021, *140* (1), 1–20.

_ , Robert J. Richmond, and Motohiro Yogo, "Which Investors Matter for Equity Valuations and Expected Returns?," *Review of Economic Studies*, 2024, *91* (4), 2387–2424.

Li, Jiacui and Zihan Lin, "Price Multipliers Are Larger for Less Diversifiable Order Flows," *Journal of Finance*, 2026. Forthcoming.

Li, Zheng, Anna Pavlova, and Taisiya Sikorskaya, "Sparse Portfolios and Benchmarking in Corporate Bond Markets," Working Paper 6350978, SSRN 2026. Posted March 2026; revised August 2026.

Lou, Dong, "A Flow-Based Explanation for Return Predictability," *Review of Financial Studies*, 2012, *25* (12), 3457–3489.

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," in Paul Zarembka, ed., *Frontiers in Econometrics*, Academic Press, 1974, pp. 105–142.

Mullahy, John, "Multivariate Fractional Regression Estimation of Econometric Share Models," *Journal of Econometric Methods*, 2015, *4* (1), 71–100.

Nevo, Aviv, "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 2001, *69* (2), 307–342.

Papke, Leslie E. and Jeffrey M. Wooldridge, "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates," *Journal of Applied Econometrics*, 1996, *11* (6), 619–632.

Pavlova, Anna and Taisiya Sikorskaya, "Benchmarking Intensity," *Review of Financial Studies*, 2023, *36* (3), 859–903.

Petajisto, Antti, "Why Do Demand Curves for Stocks Slope Down?," *Journal of Financial and Quantitative Analysis*, 2009, *44* (5), 1013–1044.

Shleifer, Andrei, "Do Demand Curves for Stocks Slope Down?," *Journal of Finance*, 1986, *41* (3), 579–590.

Silva, J. M. C. Santos and Silvana Tenreyro, "The Log of Gravity," *Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Staiger, Douglas and James H. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 1997, *65* (3), 557–586.

Sufi, Amir, "Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans," *Journal of Finance*, 2007, *62* (2), 629–668.

U.S. Securities and Exchange Commission, "Form 13F: Report of Institutional Investment Manager," Special Instruction 9 (omission of small positions); see also Division of Investment Management, "Frequently Asked Questions About Form 13F," Q.39 2025. A manager may omit a holding only if it holds fewer than 10,000 shares (or less than $200,000 principal amount of convertible debt) *and* the aggregate fair market value is less than $200,000. Both conditions must hold.

van Binsbergen, Jules H., Benjamin David, and Christian C. Opp, "How (Not) to Identify Demand Elasticities in Dynamic Asset Markets," 2026. Working paper, March 27, 2026; earlier version NBER Working Paper 34528, December 2025.

van der Beck, Philippe, "Short versus Long-Run Demand Elasticities in Asset Pricing," *Journal of Financial Economics*, 2026, p. 104337.

Wurgler, Jeffrey and Ekaterina Zhuravskaya, "Does Arbitrage Flatten Demand Curves for Stocks?," *Journal of Business*, 2002, *75* (4), 583–608.

A The Level and Ratio Moments

This appendix relates the two moments of Section 2 to familiar estimators, sets out the identifying restriction behind the level moment and its sensitivity to reporting and implementation departures, and compares the two moments when the tail of a portfolio is omitted.

A.1 The level and ratio moments as quasi-likelihood scores

For a single portfolio, the level moment (9) is the quasi-score of a conditional-logit analogue of the fractional multinomial logit (Papke and Wooldridge, 1996; Mullahy, 2015). Equivalently, it is the first-order condition of Poisson pseudo-maximum likelihood on dollar positions with a portfolio fixed effect (Guimarães et al., 2003): concentrating out the fixed effect leaves $\sum_j x_j\{w_j - \mu_j(\theta)\} = 0$, so the two estimators coincide portfolio by portfolio. The ratio moment (5) is the corresponding Gamma pseudo-likelihood score. The choice between them is a choice of variance function, the Poisson case corresponding to residual variance proportional to the fitted share and the Gamma case to variance proportional to its square, as in Santos Silva and Tenreyro (2006). These equivalences hold when the characteristics serve as their own instruments; with an excluded instrument for price, each moment is the instrumental-variables analogue of the corresponding score.

Koijen and Yogo (2019) consider a related but different choice in their Appendix F, comparing estimation "in levels with zero holdings" against a log-linear specification that drops zero holdings, and cite Santos Silva and Tenreyro (2006) for the analogous issue in gravity estimation. That choice concerns the treatment of zero holdings and the transformation of the dependent variable. The choice studied here concerns how deviations are weighted. In the variance-function taxonomy their benchmark estimator is the Gamma case, which is the ratio moment of Section 2.2, and the level moment is the Poisson case.

A.2 Identifying restriction and sensitivity

The level moment changes both the treatment of allocation discrepancies and the restriction used to identify demand. To see the distinction, write it as a fitted-share-weighted ratio moment:

$$\sum_{j\in U} z_j\{\tilde{w}_j - \mu_j(\theta)\} = \sum_{j\in U} \mu_j(\theta) z_j \left\{\frac{\tilde{w}_j}{\mu_j(\theta)} - 1\right\}. \tag{22}$$

A given relative discrepancy receives more weight in a large position than in a small one. This weighting limits the influence of relative shortfalls among small positions, but the fitted weights contain endogenous prices. The identifying restriction underlying ratio estimation therefore does not automatically justify the level moment.

Substituting the observed-holdings specification (6) gives, at the demand parameters,

$$E\left[z_j\{\tilde{w}_j - \mu_j(\theta_0)\}\right] = \underbrace{E\left[z_j\mu_j(\theta_0)(\varepsilon_j - 1)\right]}_{\text{Latent demand}} + \underbrace{E\left[z_j c_j\right]}_{\substack{\text{Reporting and}\\ \text{implementation}\\ \text{departures}}}. \tag{23}$$

These two components raise distinct questions: whether latent demand satisfies the restriction after weighting by position size, and how much reporting and implementation departures perturb the resulting moment.

For latent demand, the relevant restriction is

$$E\left[z_j \mu_j(\theta_0)(\varepsilon_j - 1)\right] = 0. \tag{24}$$

The KY restriction, $E[\varepsilon_j \mid z_j] = 1$ in (3), makes relative latent-demand departures conditionally mean zero. Level estimation additionally requires orthogonality after those departures are expressed in absolute units. Under the KY restriction,

$$E\left[z_j \mu_j(\varepsilon_j - 1)\right] = E\left[z_j \operatorname{Cov}(\mu_j, \varepsilon_j \mid z_j)\right]. \tag{25}$$

The concern arises because prices respond to demand shocks shared across investors. Even when the instrument is unrelated to the focal investor's relative demand departure, the fitted position can remain related to that departure through price.

Such dependence need not distort the estimated demand slopes. If its effect on expected demand is a common proportional rescaling across securities, the portfolio intercept can absorb it. The threat to slope identification is variation in that rescaling that aligns with the instruments or included characteristics. For example, common demand shocks could have different effects across small and large stocks, producing systematic differences in expected departures in absolute units. The instrument's exclusion restriction addresses its relationship with latent demand, but does not by itself rule out this additional channel through fitted position size.

Table 5 illustrates the difference between the two restrictions. The same relative latent-demand departure of 30 percent moves a position with a fitted weight of 7 percent by 2.1 percentage points and a position with a fitted weight of 0.01 percent by 0.003 percentage points. The KY restriction concerns the relative departures in the fourth column; the level restriction (24) concerns the departures in absolute units in the fifth. One way the level restriction can hold is for relative departures to balance, on average, among stocks with similar fitted position sizes and instrument values, as illustrated in the table. Relative departures can instead balance within an instrument group while departures in absolute units do not: positive tilts may fall systematically on larger positions and negative tilts on smaller ones. The resulting conditional covariance between fitted position size and latent demand is the additional term in (25).

Reporting and implementation departures require a separate assessment. Exact orthogonality of their contribution, $E[z_j c_j] = 0$, together with the latent-demand restriction is sufficient for the level moment to hold. Systematic omission need not satisfy this condition. The case for level estimation therefore concerns sensitivity to such departures as well as identification under exact orthogonality.

In absolute units, that sensitivity depends on the capital reallocated and the instruments of the affected securities. Moving a small amount of capital from omitted constituents to retained positions makes a small direct contribution to the level moment when the associated instrument differences are bounded. Increasing the number of omitted names does not indepen-

Table 5: **Latent-demand departures in relative and absolute units.**

Position	Fitted weight μ_j	Holding $w_j^* = \mu_j \varepsilon_j$	$\varepsilon_j - 1$	$\mu_j(\varepsilon_j - 1)$
Large	7%	9.1%	+30%	+2.1 pp
Large	7%	4.9%	−30%	−2.1 pp
		⋮		
Small	0.01%	0.013%	+30%	+0.003 pp
Small	0.01%	0.007%	−30%	−0.003 pp

Notes. Hypothetical positions, without reporting or implementation departures ($c_j = 0$). The KY restriction (3) requires the relative departures $\varepsilon_j - 1$ to be conditionally mean zero given the instruments. The level moment additionally requires the departures in absolute units, $\mu_j(\varepsilon_j - 1)$, to be orthogonal to the instruments, restriction (24).

dently increase this contribution if their total weight remains small. Appendix A.3 formalizes this distinction between the number of affected securities and their share of portfolio capital. The resulting effect on estimated coefficients also depends on the strength of identification. Estimation in levels limits the direct influence of small allocation departures; it does not require treating systematic omission as exactly mean zero.

To assess the importance of the additional latent-demand restriction, I also estimate a version using weights constructed from the instruments. Let $\tilde{\mu}_j(\theta)$ denote the fitted allocation obtained by replacing log market capitalization in the demand index with its first-stage prediction from the instruments and included characteristics. The predetermined-weight estimator solves[14]

$$\sum_{j \in U} z_j \, \tilde{\mu}_j(\hat{\theta}) \left\{ \frac{\tilde{w}_j}{\mu_j(\hat{\theta})} - 1 \right\} = 0. \tag{26}$$

For target demand, weighting by a function of the instruments preserves the KY restriction: $E[h(z_j)(\varepsilon_j - 1)] = 0$. Applied to observed holdings, measurement error contributes

$$E\left[z_j c_j \frac{\tilde{\mu}_j}{\mu_j} \right]. \tag{27}$$

When predicted and fitted allocations are close, this contribution is approximately in absolute units. The comparison asks whether weighting relative discrepancies by predicted position size reproduces the level estimates while retaining the KY restriction on latent demand.

On the common estimation sample, the asset-weighted mean price coefficient is 0.707 under the level moment, 0.710 under the predetermined-weight moment, and 0.811 under the ratio moment.[15] For broad index trackers, whose benchmark is one, the corresponding estimates are 1.016, 1.018, and 1.094. The predetermined-weight estimates thus reproduce nearly all of the movement from ratio to level estimation, supporting the interpretation that the weighting of allocation discrepancies accounts for most of the empirical difference.

The predetermined-weight estimator does not converge in every fund-quarter for which

[14] The first stage is estimated separately for each fund-quarter, regressing $\log \mathrm{ME}_j$ on z_j across the fund's investment universe. Because $\tilde{\mu}_j$ depends on θ, I evaluate it at the estimate itself and solve the resulting fixed point, starting from the level estimate.

[15] Size-band index funds, which follow capitalization only within a size range, are the exception. Excluding them, the three estimates are 0.709, 0.710 and 0.795.

the level estimator has a solution, and I use it as a robustness check. Its proximity to the level estimates provides evidence that the additional latent-demand restriction has limited influence on the reported results.

A.3 Omitted positions under the two moments

Proposition 1 (Omitted tail, ratio moment). *Suppose latent demand is constant, a single characteristic x_j serves as its own instrument, $z_j = (1, x_j)'$, and the portfolio leaves out a set $\mathcal{T}$ of $N_\mathcal{T}$ securities with total model weight $m_\mathcal{T} = \sum_{j\in\mathcal{T}} \mu_j(\theta_0)$, scaling up the rest proportionally. Then, to first order, the ratio-moment estimate of the coefficient on x_j has bias*

$$\text{Bias}^{\text{rat}} \approx -\frac{N_\mathcal{T}}{1-m_\mathcal{T}} \frac{\bar{x}_\mathcal{T} - \bar{x}}{\sum_j (x_j - \bar{x})^2},$$

where $\bar{x}$ and $\bar{x}_\mathcal{T}$ are the unweighted means of x_j over U and over $\mathcal{T}$.

Proof. Write $\mu_j = \mu_j(\theta_0)$, with $\mu_j(\theta) = \exp(a + bx_j)$ as in Section 2.1, so that $\sum_{j\in U} \mu_j = 1$ and $\partial\mu_j/\partial\theta' = \mu_j z_j'$ for $\theta = (a, b)'$. The observed weights are $\tilde{w}_j = \mu_j/(1 - m_\mathcal{T})$ for $j \notin \mathcal{T}$ and $\tilde{w}_j = 0$ for $j \in \mathcal{T}$. For a moment $g(\theta)$, to first order $\hat{\theta} - \theta_0 \approx -G^{-1} g(\theta_0)$, where $G = \partial g/\partial\theta'$ is evaluated at the portfolio without omission. For the ratio moment $g(\theta) = \sum_j z_j\{\tilde{w}_j/\mu_j(\theta) - 1\}$, the displacement at θ_0 is $g(\theta_0) = \sum_j z_j d_j$, with $d_j = m_\mathcal{T}/(1 - m_\mathcal{T})$ for held names and $d_j = -1$ for omitted names, and the sensitivity is $G = -\sum_j z_j z_j'$. Hence $\hat{\theta} - \theta_0 \approx (\sum_j z_j z_j')^{-1} \sum_j z_j d_j$, the least-squares coefficients of d_j on $(1, x_j)$ across U, whose slope is $\sum_j (x_j - \bar{x}) d_j / \sum_j (x_j - \bar{x})^2$. Because $\sum_{j\in U}(x_j - \bar{x}) = 0$, the held names satisfy $\sum_{j\notin\mathcal{T}}(x_j - \bar{x}) = -N_\mathcal{T}(\bar{x}_\mathcal{T} - \bar{x})$, so $\sum_j (x_j - \bar{x}) d_j = -N_\mathcal{T}(\bar{x}_\mathcal{T} - \bar{x})\{m_\mathcal{T}/(1 - m_\mathcal{T}) + 1\} = -N_\mathcal{T}(\bar{x}_\mathcal{T} - \bar{x})/(1 - m_\mathcal{T})$. □

Because $\sum_j (x_j - \bar{x})^2$ grows with the number of securities, the bias scales with the share of names in the omitted tail rather than with the capital they carry, and it is positive when the omitted names are small.

Proposition 2 (Omitted tail, level moment). *Under the conditions of Proposition 1, with x_j normalized to have model-weighted mean zero and model-weighted second moment one, the level-moment estimate of the coefficient on x_j has first-order bias*

$$\text{Bias}^{\text{lev}} \approx -\frac{m_\mathcal{T}}{1-m_\mathcal{T}} \bar{x}^\mu_\mathcal{T},$$

where $\bar{x}^\mu_\mathcal{T}$ is the model-weighted mean of x_j over $\mathcal{T}$.

Proof. With the notation of the previous proof, the level moment is $g(\theta) = \sum_j z_j\{\tilde{w}_j - \mu_j(\theta)\}$. Its displacement at θ_0 is $g(\theta_0) = \sum_j \mu_j z_j d_j$, with the same d_j, and its sensitivity is $G = -\sum_j \mu_j z_j z_j'$. Hence $\hat{\theta} - \theta_0 \approx (\sum_j \mu_j z_j z_j')^{-1} \sum_j \mu_j z_j d_j$, the coefficients of a regression of d_j on $(1, x_j)$ weighted by μ_j. Under the normalization, $\sum_j \mu_j z_j z_j' = I$, so the slope is $\sum_j \mu_j x_j d_j$. Because $\sum_{j\in U} \mu_j x_j = 0$, the held names satisfy $\sum_{j\notin\mathcal{T}} \mu_j x_j = -m_\mathcal{T} \bar{x}^\mu_\mathcal{T}$, so $\sum_j \mu_j x_j d_j = -m_\mathcal{T} \bar{x}^\mu_\mathcal{T}\{m_\mathcal{T}/(1 - m_\mathcal{T}) + 1\} = -m_\mathcal{T} \bar{x}^\mu_\mathcal{T}/(1 - m_\mathcal{T})$. □

Under the ratio moment the omitted-tail bias scales with the number of names in the tail; under the level moment it scales with the tail's share of capital. In equity portfolios the tail holds many of the names and little of the capital (Section 1.1).

B Aggregation of Heterogeneous Mandates

This appendix derives the consolidated demand index used in Section 2.4. It shows how portfolio normalization determines the weight each constituent mandate receives, writes the consolidated index as a cumulant generating function, and characterizes the slope and curvature that a linear specification fitted to consolidated holdings must approximate.

B.1 Consolidated holdings and mixing weights

Consider a portfolio formed by combining mandates $m = 1, \ldots, M$, such as the funds of one fund family, with capital shares $a_m \geq 0$ and $\sum_m a_m = 1$. Write $p_j \equiv \log \mathrm{ME}_j$. Mandate m follows the demand system of Section 2.1 with its own coefficients,

$$w_{jm} = \frac{e^{\delta_{jm}}}{D_m}, \qquad \delta_{jm} = \alpha_m p_j + X_j' \beta_m, \qquad D_m = \sum_k e^{\delta_{km}}, \tag{28}$$

where D_m is the mandate's normalizer: the sum runs over its investment universe, and includes a unit term when the mandate holds an outside option.[16] The consolidated portfolio holds $W_j = \sum_m a_m w_{jm}$, so

$$W_j = C \sum_m \pi_m e^{\delta_{jm}}, \qquad \pi_m \equiv \frac{a_m / D_m}{\sum_\ell a_\ell / D_\ell}, \qquad C \equiv \sum_\ell \frac{a_\ell}{D_\ell}. \tag{29}$$

The constant C is common to all securities and is absorbed by the portfolio intercept or by renormalization within the universe. With an outside option, $1/D_m = w_{0m}$ and $\pi_m = a_m w_{0m} / \sum_\ell a_\ell w_{0\ell}$.

The mixing weights are not the capital shares:

$$\pi_m \neq a_m \quad \text{unless all mandates have the same normalizer } D_m. \tag{30}$$

Each mandate enters deflated by the scale of its own demand index over its universe. Two mandates with the same capital can therefore receive materially different weight in the consolidated portfolio if they differ in the breadth of their universes, in how much of their capital they hold outside it, or in the level of their indices. Because D_m depends on prices, the mixing weights also move with prices.

[16] To isolate heterogeneity in the systematic indices, latent demand is held constant. A latent component common to all mandates adds to every δ_{jm} and passes through the consolidation unchanged. A security outside a mandate's universe receives no weight from that mandate.

B.2 The cumulant generating function representation

Taking logs in (29) and dropping the common constant gives the consolidated demand index

$$\Lambda_j \equiv \log \sum_m \pi_m e^{\delta_{jm}} = \log E_\pi \left[e^{\delta_{jm}} \right]. \tag{31}$$

Result 1 (CGF representation)**.** Let Δ_j take the value δ_{jm} with probability π_m. The consolidated index (31) is the cumulant generating function of Δ_j evaluated at one. Its cumulant expansion is

$$\Lambda_j = E_\pi[\delta_{jm}] + \tfrac{1}{2} \operatorname{Var}_\pi(\delta_{jm}) + \tfrac{1}{6} c_3(\delta_{jm}) + \cdots, \tag{32}$$

where c_3 is the third cumulant under π.

The first term is a demand index with π-weighted average coefficients. The remaining terms are zero only when the mandates' indices coincide for security j. Consolidation therefore presents the estimator with the dispersion of allocation rules as well as their average.

B.3 Slope and curvature

Holding the mixing weights fixed, differentiating (31) gives

$$\frac{\partial \Lambda_j}{\partial p_j} = \sum_m \tilde{\pi}_{jm} \alpha_m, \qquad \frac{\partial^2 \Lambda_j}{\partial p_j^2} = \operatorname{Var}_{\tilde{\pi}_j}(\alpha_m) \geq 0, \qquad \tilde{\pi}_{jm} \equiv \frac{\pi_m e^{\delta_{jm}}}{\sum_\ell \pi_\ell e^{\delta_{j\ell}}}. \tag{33}$$

The tilted weights $\tilde{\pi}_{jm}$ are the shares of mandate m in the consolidated holding of security j. A higher price raises the relative influence of mandates with larger price coefficients, so the slope of consolidated demand rises with price even though every constituent index is linear. The curvature is zero only if the mandates share a common price coefficient.

Allowing the normalizers to respond to the price changes the expressions only by terms in the mandates' own weights in security j. For the log of the observed consolidated weight, the exact derivatives are

$$\frac{d \log W_j}{d p_j} = \sum_m \tilde{\pi}_{jm} \alpha_m (1 - w_{jm}), \qquad \frac{d^2 \log W_j}{d p_j^2} = \operatorname{Var}_{\tilde{\pi}_j}\big(\alpha_m (1 - w_{jm})\big) - \sum_m \tilde{\pi}_{jm} \alpha_m^2 w_{jm} (1 - w_{jm}), \tag{34}$$

which reduce to (33) when security j is a small position in each mandate. Because $\tilde{\pi}_{jm} = a_m w_{jm} / W_j$ is mandate m's share of the consolidated holding of security j, the first expression implies that the consolidated quantity elasticity is

$$1 - \frac{d \log W_j}{d p_j} = \sum_m \tilde{\pi}_{jm} \big\{ 1 - \alpha_m (1 - w_{jm}) \big\}, \tag{35}$$

the ownership-weighted average of the mandates' elasticities, as in (16). This is the object that Section 3 constructs by estimating mandates separately and aggregating afterward. It varies across securities with the composition of their holders, which a single linear index fitted to consolidated holdings cannot represent.

B.4 Fitting a linear index to consolidated holdings

To second order, (32) gives

$$\Lambda_j \approx \bar{\alpha}\, p_j + X_j'\bar{\beta} + \Gamma_j, \qquad \Gamma_j \equiv \tfrac{1}{2}\operatorname{Var}_\pi(\alpha_m)p_j^2 + p_j X_j' \operatorname{Cov}_\pi(\alpha_m, \beta_m) + \tfrac{1}{2} X_j' \operatorname{Var}_\pi(\beta_m) X_j, \quad (36)$$

where $\bar{\alpha} = E_\pi[\alpha_m]$ and $\bar{\beta} = E_\pi[\beta_m]$. The term Γ_j is not noise: it is a deterministic function of the security's characteristics implied by heterogeneity across mandates.

Corollary 1 (The consolidated coefficient)**.** *Suppose a linear index $\Lambda_j = c + \alpha p_j + X_j'\beta +$ residual is fitted by least squares, and let $\tilde{p}_j$ be the residual from projecting p_j on a constant and X_j. Then, to second order,*

$$\hat{\alpha} - \sum_m a_m \alpha_m \approx \underbrace{\sum_m (\pi_m - a_m)\alpha_m}_{\textit{composition}} + \underbrace{\frac{\operatorname{Cov}(\tilde{p}_j, \Gamma_j)}{\operatorname{Var}(\tilde{p}_j)}}_{\textit{curvature}}, \quad (37)$$

where the covariance and variance are taken across securities. The fitted residual contains Γ_j net of its projection on $(1, p_j, X_j)$.

The consolidated coefficient therefore differs from the capital-weighted average of the constituent coefficients for two reasons. The composition term arises because normalization weights mandates by π_m rather than by their capital. The curvature term arises because a single slope is fitted to a convex index; its sign depends on how residualized price covaries with Γ_j across securities, so curvature can move the coefficient in either direction. The part of Γ_j that the linear index cannot represent appears in the residual, as apparent latent demand. Both terms are properties of the demand object presented to the estimator, so neither disappears as the number of securities grows. Least squares is used here for transparency; the moment estimators of Section 2 fit the same convex object with a single slope and face the same two terms with different weights across securities.

C Data, Specification, and Implementation Details

This appendix collects the data construction and implementation details that are deliberately suppressed in Section 3, where the main text focuses on the three principles of the estimation design: the moment condition, the observation unit, and the level at which estimates and their uncertainty are aggregated.

C.1 Data and observation-unit construction

To isolate the estimator and observational-unit issues studied in Section 2, the implementation stays as close as possible to the canonical demand-system setup of Koijen and Yogo (2019). The main departures are the two that Section 2 motivates: level rather than relative-error moments, and estimation at the mandate rather than the consolidated-institution level. The characteristic specification and identification strategy are kept as close as possible to the canonical

implementation. They are not identical: the investor population, the holdings source and the conditioning on the equity universe differ, and the instrument is adapted to the mutual-fund universe.

Stock characteristics follow Koijen and Yogo (2019) in spirit. Accounting variables enter from six months after the fiscal period end and remain in use for at most 18 months after it. The dividend characteristics are a payer indicator and the magnitude of trailing twelve-month cash dividends relative to lagged book equity. Investment is the simple growth of total assets, and market beta is estimated over 36 months.

The central change is the observational unit of holdings. Canonical applications often rely on consolidated institutional portfolios from SEC 13F filings, where the observed "institution" is typically a fund family or manager that aggregates multiple underlying mandates. In contrast, I use the CRSP Mutual Fund Database and treat each mutual fund portfolio as a mandate-like decision unit. Conceptually, this shift moves the data closer to the primitive decision maker and removes aggregation across heterogeneous sleeves by construction. Empirically, it replaces the institution-level rollup with fund-level holdings that are still large and diffuse, but substantially closer to a single investment style than a 13F filer.

I restrict attention to funds that Lipper classifies as equity (asset code `EQ`) and use their positions in U.S. common stocks listed on the NYSE, AMEX and Nasdaq.[17] The asset code admits international, global and balanced funds, whose weights are normalized over their U.S. common stock positions, so the demand coefficients reflect reallocation within those positions rather than cross-asset allocation decisions. Where the analysis groups funds by style or index status, it uses point-in-time Lipper classes and CRSP index flags.

The CRSP Mutual Fund Database reports holdings at the share-class level, where multiple share classes (e.g., retail vs. institutional) may correspond to the same underlying portfolio but differ in fee structure, distribution channel, and investor eligibility. For the purposes of demand estimation, these share classes implement the same portfolio decision. I therefore aggregate share classes to the portfolio level (CRSP portfolio number) and compute portfolio AUM as the sum of AUM across share classes within the same portfolio. Holdings are correspondingly treated as portfolio holdings. This aggregation reduces mechanical duplication, aligns the observational unit with the portfolio choice, and avoids overweighting portfolios that simply offer many share classes.

I construct each fund's effective universe to mirror the benchmark demand-system setup. In each quarter t, the universe for a fund is defined as the union of securities observed in the fund's holdings at t and the securities that appeared in the fund's holdings over the preceding eleven quarters. This rolling-union universe matches the intuition that portfolio choice is made over a persistent opportunity set and ensures that the estimator does not mechanically treat previously held securities as unavailable options. It also maintains continuity with the large-universe logic in Koijen and Yogo (2019): funds choose from a broad set of securities, and the econometrician must account for the fact that most weights are small.

[17] CRSP leaves the permanent stock identifier blank on some holdings of U.S. common stocks and reports the firm's later CUSIP. I relink these rows through the CUSIP on the stock header, or otherwise the historical CUSIP, when the resulting stock is in the stock file for that quarter. They are 2.08% of holdings market value in 2010Q2–2016Q4, 1.21% in 2017–2018 and 0.38% in 2019–2024Q3.

Table 6: Mutual-fund sector coverage and ownership (2010–2024)

	# Funds	Mkt cap	% Market held (MF sector)					Total AUM
Year		($bn)	All	Large	Mid	Small	Micro	($bn)
2010	7,070	13,169.3	24.5	23.7	31.1	29.8	16.5	7,217.2
2011	7,453	14,370.8	26.7	25.7	34.0	34.0	19.4	8,000.5
2012	7,752	15,639.6	26.7	25.8	33.7	33.5	18.9	8,665.0
2013	7,865	18,704.2	27.1	26.1	33.6	34.6	18.5	10,490.9
2014	8,182	21,745.3	27.6	26.6	34.2	34.9	18.1	12,377.7
2015	8,541	21,882.9	27.7	26.7	35.0	34.1	16.6	12,684.2
2016	8,782	22,223.0	28.3	27.4	36.7	34.7	17.2	13,116.0
2017	8,736	25,649.1	28.6	27.7	36.9	35.4	17.3	15,370.3
2018	8,754	27,506.4	29.1	28.2	37.4	36.8	17.2	16,192.5
2019	8,751	29,627.0	29.5	28.7	38.6	36.5	16.1	17,433.1
2020	8,584	32,191.8	29.1	28.5	37.1	34.1	13.7	17,817.3
2021	8,540	45,428.5	28.5	28.1	35.4	32.6	10.7	23,545.4
2022	8,807	39,561.2	28.5	28.0	37.7	30.4	7.9	20,665.3
2023	8,929	43,310.4	28.8	28.2	38.6	33.4	11.8	22,213.2
2024	9,007	54,291.4	29.0	28.4	38.8	34.8	14.6	26,798.0

Notes: Entries are annual averages across quarters within each year. "% Market held" is mutual-fund dollar holdings divided by total market capitalization in the corresponding stock universe. Size buckets follow the paper's size breakpoints. The statistics cover all CRSP mutual funds, not the estimation sample (Section 3.1).

Table 7: Mutual-fund sector coverage: portfolio granularity and tail concentration (2010–2024)

	AUM/fund	Universe size		# Holdings		$w < 1\%$		$w < 0.1\%$	
Year	Mean($mn)	Mean	P90	Mean	P90	% Hld	% Mass	% Hld	% Mass
2010	1,020.6	358	662	216	460	54.7	33.3	16.1	4.13
2011	1,074.5	370	682	219	450	55.1	33.6	16.4	4.25
2012	1,117.8	381	700	223	450	55.0	32.8	17.3	4.42
2013	1,333.2	411	734	236	485	55.3	34.4	17.6	4.78
2014	1,512.7	432	810	244	545	55.5	34.9	18.1	4.93
2015	1,485.9	445	924	252	574	56.8	35.6	19.2	5.14
2016	1,493.6	462	1,157	263	698	57.8	36.1	20.7	5.43
2017	1,759.7	481	1,288	274	881	58.4	36.8	21.5	5.71
2018	1,849.9	499	1,584	280	1,035	58.7	36.6	22.9	6.02
2019	1,992.4	539	1,980	298	1,129	62.1	38.3	25.1	6.47
2020	2,077.4	536	1,993	291	1,083	62.7	37.3	26.5	6.21
2021	2,756.6	592	2,157	318	1,187	63.8	39.1	27.4	6.98
2022	2,347.9	618	2,351	313	1,127	64.4	39.3	27.7	6.81
2023	2,487.6	615	2,592	315	1,097	65.0	38.6	28.2	6.74
2024	2,974.9	615	2,610	317	1,034	67.3	38.1	30.8	6.99

Notes: Entries are annual averages across quarters within each year. Universe size is the rolling-union universe (current quarter plus prior 11 quarters); holdings counts are within the same universe. Mean and P90 for universe size, holdings, and concentration columns are AUM-weighted averages across funds. "% Hld" is the fraction of holdings with weight below the indicated threshold; "% Mass" is the total portfolio weight in those positions. The statistics cover all CRSP mutual funds, not the estimation sample (Section 3.1).

A limitation of the mutual-fund holdings data is coverage. The CRSP mutual fund holdings panel is available from 2000 onward, but coverage in the early years is thin. The reported analysis covers 2010Q2–2024Q4 (59 quarters), when mutual-fund coverage of the equity market is dense enough to support it. The sample starts in 2010Q2 because of a coverage break in the source data: about 2,200 existing CRSP funds have no reported holdings until that quarter.

Tables 6 and 7 summarize the mutual-fund sector; they cover all CRSP mutual funds, not the estimation sample. Table 6 covers the mutual-fund sector from 2010–2024, reporting the number of funds, aggregate AUM, and the fraction of market capitalization held by mutual funds overall and by size bin. Table 7 reports portfolio granularity and tail-concentration statistics: AUM per fund, rolling-universe size and number of holdings (AUM-weighted mean and P90), and the fraction of holdings with weight below 1% and 0.1% together with the portfolio mass those positions represent. Its long-tail features motivate the remainder of Section 3: by 2024 an average fund holds over 300 names in a universe of roughly 600 securities, with 67% of positions carrying weight below 1% of the portfolio.

C.2 Specification and interpretation

Variables and characteristics. Let f index a mutual fund, j index stocks, and t index quarters. For each fund-quarter (f,t), let $\mathcal{U}_{ft}$ denote the fund's effective universe and let $w_{fjt}^{\text{raw}} \geq 0$ be the reported portfolio weight for stock $j \in \mathcal{U}_{ft}$, as a share of the fund's total assets. Define the within-universe total

$$S_{ft} \equiv \sum_{j \in \mathcal{U}_{ft}} w_{fjt}^{\text{raw}}. \tag{38}$$

When funds hold cash or assets outside $\mathcal{U}_{ft}$, S_{ft} need not equal one. The weights used throughout are the normalized weights $w_{fjt} \equiv w_{fjt}^{\text{raw}} / S_{ft}$, so that $\sum_{j \in \mathcal{U}_{ft}} w_{fjt} = 1$. The estimator thus identifies demand parameters from relative allocations within $\mathcal{U}_{ft}$.

The characteristics, motivated by the Fama–French five-factor model, are $\log(\text{B/M})_{jt}$ and the vector X_{jt} of the remaining attributes:

- $\log(\text{B/M})_{jt}$: log book-to-market ratio;
- OP_{jt}: operating profitability (revenue minus cost of goods sold, SGA, and interest, divided by lagged book equity, annualized over a trailing four-quarter window);
- Inv_{jt}: asset growth (change in total assets over four quarters, scaled by lagged total assets);
- $\beta_{j,t-1}^{mkt}$: market beta estimated from a trailing 36-month rolling regression of excess stock returns on excess market returns;
- I_{jt}^{div}: a dividend-payer indicator (equals one if the firm paid a positive dividend within the trailing twelve months); and
- $\log(1 + D/\text{BE})_{jt}$: log dividend magnitude among payers (dividends to book equity, set to zero for non-payers).

All continuous characteristics are winsorized at the 2.5th and 97.5th percentiles each quarter and then standardized to zero mean and unit variance. The dividend specification decomposes the extensive margin (payer vs. non-payer) from the intensive margin (how much, conditional on paying), following the observation that a large fraction of stocks pay no dividends and a

single dividend variable would be dominated by the mass point at zero.

The demand specification. My specification follows the relative-demand component of Koijen and Yogo (2019)'s equation (10). For each (f,t) and $j \in \mathcal{U}_{ft}$, define a linear index

$$\eta_{fjt}(\theta_{ft}) \;=\; \alpha_{ft}\,\log \mathrm{ME}_{jt} \;+\; \beta_{ft}^{BM}\,\log(\mathrm{B/M})_{jt} \;+\; X_{jt}'\,\beta_{ft}, \tag{39}$$

where $\theta_{ft} \equiv (\alpha_{ft}, \beta_{ft}^{BM}, \beta_{ft})$ are the parameters of interest, and the implied within-equity share

$$\mu_{fjt}(\theta_{ft}) \;=\; \frac{\exp(\eta_{fjt})}{\sum_{k\in\mathcal{U}_{ft}} \exp(\eta_{fkt})}, \tag{40}$$

so that $\sum_{j\in\mathcal{U}_{ft}} \mu_{fjt}(\theta_{ft}) = 1$ for all θ_{ft}. The residual is $u_{fjt}(\theta_{ft}) = w_{fjt} - \mu_{fjt}(\theta_{ft})$.

The object estimated here is therefore *within-equity demand*: given the fund's allocation to the equity universe, how does it allocate across stocks? The main implementation difference from the benchmark outside-option specification is that I do not estimate a free intercept. In Koijen and Yogo (2019), the intercept governs the level of inside-equity demand relative to the outside option and is identified through the residual normalization. Once holdings are normalized within the equity universe, any common level shift cancels from the logit (equation (42)) and is not identified from within-equity share moments. The next paragraph derives the exact mapping from the outside-option formulation to the within-equity specification used here.

Scope of identification: within-equity demand. It is worth making explicit the relation between this within-equity specification and the outside-option formulation in Koijen and Yogo (2019). Write Koijen and Yogo (2019)'s equation (10) compactly as

$$\frac{w_{fj}}{w_{f0}} = \exp\{\beta_f^0 + r_j'\theta_f\}\,\varepsilon_{fj},$$

where β_f^0 is the intercept ($\beta_{K,i}$ in their notation) and $r_j'\theta_f$ collects the nonconstant characteristics. The budget constraint over the outside option and the inside assets implies

$$w_{fj} = \frac{\exp\{\beta_f^0 + r_j'\theta_f\}\,\varepsilon_{fj}}{1 + \sum_{k\in\mathcal{N}_f} \exp\{\beta_f^0 + r_k'\theta_f\}\,\varepsilon_{fk}}, \quad w_{f0} = \frac{1}{1 + \sum_{k\in\mathcal{N}_f} \exp\{\beta_f^0 + r_k'\theta_f\}\,\varepsilon_{fk}}.$$

The intercept β_f^0 shifts total inside demand relative to the outside option. It is not identified from relative inside shares; it is identified by the outside-option odds together with the residual normalization $\mathbb{E}[\varepsilon_{fj} \mid z_{fj}] = 1$.

Conditioning on the inside equity universe eliminates this margin algebraically. The latent residuals enter the conditional shares unchanged, while β_f^0 cancels exactly:

$$\frac{w_{fj}}{\sum_{\ell\in\mathcal{N}_f} w_{f\ell}} = \frac{\exp\{r_j'\theta_f\}\,\varepsilon_{fj}}{\sum_{\ell\in\mathcal{N}_f} \exp\{r_\ell'\theta_f\}\,\varepsilon_{f\ell}}. \tag{41}$$

The conditional equity shares therefore identify the relative slopes θ_f, but not the equity-vs-outside intercept β_f^0. The model-implied conditional shares $\mu_{fjt}(\theta_{ft})$ in equation (40) correspond to the same expression evaluated at $\varepsilon_{fj} \equiv 1$, so the residual $u_{fjt} = w_{fjt} - \mu_{fjt}$ inherits the latent residual variation.

The distinction is conceptual rather than merely numerical. A free intercept is meaningful in the Koijen and Yogo (2019) model because the left-hand side is w_{fj}/w_{f0} and the outside option provides a level against which inside equity demand is measured. Once the data are normalized within the equity universe, the same intercept is not identified by share moments: for any scalar a,

$$\frac{\exp\{a + r_{jt}'\theta\}}{\sum_k \exp\{a + r_{kt}'\theta\}} = \frac{\exp\{r_{jt}'\theta\}}{\sum_k \exp\{r_{kt}'\theta\}}. \tag{42}$$

Because the paper's results concern within-equity slopes, I do not reintroduce a Koijen and Yogo (2019)-style intercept, which would require either an additional outside-option measurement or an auxiliary normalization on the latent residual.

Level versus ratio moments. The level moment is (15). The moment used in the canonical implementation of Koijen and Yogo (2019) is the orthogonality condition $\mathbb{E}[z_{fjt}(\varepsilon_{fjt} - 1)] = 0$ imposed on the multiplicative latent demand ε_{fjt} of their share equation, normalized so that $\mathbb{E}[\varepsilon_{fjt} \mid z_{fjt}] = 1$ (see the scope-of-identification paragraph above). In that specification the share equation is written relative to an outside option, which makes the observed share $w_{fjt} = \mu_{fjt}(\theta)\,\varepsilon_{fjt}$ position by position, so the sample analogue is $\sum_j z_{fjt}\{w_{fjt}/\mu_{fjt}(\theta) - 1\} = 0$, a ratio moment that can be written as

$$g_{ft}^{\text{rat}}(\theta) \equiv \sum_{j \in \mathcal{U}_{ft}} z_{fjt} \left\{ \frac{w_{fjt}}{\mu_{fjt}(\theta)} - 1 \right\} = \sum_{j \in \mathcal{U}_{ft}} \frac{z_{fjt}}{\mu_{fjt}(\theta)} \{w_{fjt} - \mu_{fjt}(\theta)\} = 0. \tag{43}$$

Thus the difference is a position weighting. The level moment matches observed and fitted portfolio exposures and lets the simplex naturally dampen tiny positions. The ratio moment divides out that damping, so a one-basis-point holding can enter with the same first-order weight as a one-percent holding.

The two displays are not two moments for one estimating equation, and it is worth being explicit about which model each belongs to. Equation (43) is the moment of the canonical outside-option specification, in which $w_{fjt} = \mu_{fjt}\varepsilon_{fjt}$ holds position by position. Conditioning on the equity universe, as the implementation here does, replaces that relation by $w_{fjt} = \mu_{fjt}\varepsilon_{fjt}/\sum_k \mu_{fkt}\varepsilon_{fkt}$ (equation (41)), so w_{fjt}/μ_{fjt} recovers ε_{fjt} only up to a share-weighted normalizer common to the portfolio. The comparison drawn here does not turn on that distinction: what matters is that the ratio weighting carries the factor $1/\mu_{fjt}$ and the level weighting does not.

Price and portfolio style. The normalized exponential allocation rule in Section 2 is the within-equity share representation of the demand system used in the empirical implementation. Section 2 writes the characteristics compactly as the vector X_j; here I set out the characteristics it contains and how their coefficients are interpreted.

The specification has a natural interpretation for mutual funds because its characteristics span familiar dimensions of portfolio style. Market equity provides the price-like dimension, book-to-market the value–growth dimension, and the remaining characteristics capture other systematic tastes such as profitability and investment.

Market equity is the product of price and shares outstanding. For a local change in the stock's own price with shares outstanding fixed, $d \log \mathrm{ME}_{jt} = d \log P_{jt}$. The coefficient on log market equity can therefore be interpreted as a price coefficient once the other demand characteristics are held fixed. In this sense, α_{ft} is the mandate's *conditional price slope*: it asks how the mandate reallocates along the price-like dimension of the cross-section, net of its tastes for value, profitability, investment, and the remaining included characteristics.

The benchmark makes this interpretation transparent. A mechanically value-weighted mandate has $\alpha_{ft} = 1$ when the remaining characteristic coefficients are zero: a one-percent increase in a stock's price mechanically raises its portfolio weight by one percent.[18] Values below this benchmark imply that the mandate offsets part of that mechanical increase by reducing its relative quantity allocation as price rises. Section 3.1 translates this coefficient into the exact stock-specific quantity elasticity.

Book-to-market gives the same price movement a second economic dimension. Holding book equity fixed, a price increase raises market equity and lowers book-to-market. A growth-oriented mandate ($\beta^{BM}_{ft} < 0$) therefore finds the stock more attractive along its value–growth dimension as its price rises, whereas a value-oriented mandate ($\beta^{BM}_{ft} > 0$) finds it less attractive. The specification separates this induced style response from the conditional price slope. This distinction is useful in mutual-fund data, where size and value tilts are economically familiar features of the mandate rather than residual nuisances.

A literal price counterfactual holding book equity fixed therefore combines the two margins. Since $\log(\mathrm{B/M}) = \log \mathrm{BE} - \log \mathrm{ME}$, its slope in the systematic index is

$$\alpha^{*}_{ft} = \alpha_{ft} - \beta^{BM}_{ft}.$$

I use α_{ft} to describe the conditional price response and report the fixed-book-equity counterfactual separately when that is the relevant economic experiment. The algebraic change of basis used in estimation, and the mapping between the two coefficients, are given in Appendix C.3.

C.3 Implementation

This section collects the practical details of the estimation procedure. The baseline specification (39)–(40) defines the model; here I describe how it is taken to data.

Regressor implementation: $\log$ ME $+ \log$ BE. The specification (39) uses $\log \mathrm{ME}$ and $\log(\mathrm{B/M})$ as the two size-related regressors. In implementation, I replace $\log(\mathrm{B/M})$ with $\log \mathrm{BE}$ and es-

[18] This interpretation is exact in the frictionless mechanical benchmark where the fund does not adjust quantities when prices move and the relevant universe is fixed. For style and segmented indices (e.g., mid-cap or small-cap mandates), constituent changes and reconstitution can generate quantity adjustments when firms migrate across bins, so α need not equal one even for "passive" products. In these cases α should still be close to one if most variation is within-bin price movement rather than membership changes.

timate the index as

$$\eta_{fjt} \;=\; \alpha^{*}_{ft}\,\log \mathrm{ME}_{jt} \;+\; \beta^{BE}_{ft}\,\log \mathrm{BE}_{jt} \;+\; X'_{jt}\,\beta_{ft}, \tag{44}$$

where the asterisk on α^{*} distinguishes the coefficient in this basis from the conceptual α in (39). Since $\log(\mathrm{B/M}) = \log \mathrm{BE} - \log \mathrm{ME}$, the two specifications are algebraically equivalent, with the identification

$$\alpha_{ft} = \alpha^{*}_{ft} + \beta^{BE}_{ft}, \qquad \beta^{BM}_{ft} = \beta^{BE}_{ft}. \tag{45}$$

The practical advantage of the $\{\log \mathrm{ME}, \log \mathrm{BE}\}$ basis is that $\log \mathrm{BE}$ is predetermined and not mechanically linked to the endogenous price regressor $\log \mathrm{ME}$. This simplifies instrumentation: $\log \mathrm{BE}$ serves as its own instrument, whereas $\log(\mathrm{B/M})$ would inherit the endogeneity of $\log \mathrm{ME}$ and require separate instrumentation. I report all results in the interpretable basis (α, β^{BM}) using the conversion (45).

In implementation, all regressors are standardized to zero mean and unit variance each quarter. The price coefficient in raw (log-dollar) units is

$$\hat{\alpha}_{ft} \;=\; \frac{\hat{\alpha}^{*,\mathrm{std}}_{ft}}{\sigma_{\log \mathrm{ME}}} \;+\; \frac{\hat{\beta}^{BE,\mathrm{std}}_{ft}}{\sigma_{\log \mathrm{BE}}}, \tag{46}$$

with standard error from the delta method.[19]

Instrumental variables. Because portfolio weights w_{fjt} and stock prices p_{jt} are jointly determined in equilibrium, α is subject to the standard simultaneity concern. Following Koijen and Yogo (2019), I instrument for $\log \mathrm{ME}_{jt}$ using a measure of the stock's common fund flow pressure:

$$\mathrm{cf}^{\mathrm{me}}_{jt} \;=\; \sum_{f:\, j \in \mathcal{U}_{ft}} \frac{A_{ft}}{1 + |\mathcal{U}_{ft}|}, \tag{47}$$

where A_{ft} is fund f's total AUM. The instrument is $z_{jt} = \log(\mathrm{cf}^{\mathrm{me}}_{jt})$, standardized each quarter. To address the reflection problem, I use a leave-one-out construction: for each fund-stock pair (f, j), the instrument z^{loo}_{fjt} is the analogue of z_{jt} computed after excluding fund f's contribution to the common-flow sum in (47).[20] The moment conditions are $\mathbb{E}[z^{\mathrm{loo}}_{fjt}\,(w_{fjt} - \mu_{fjt}(\theta_{ft}))] = 0$, with all exogenous characteristics serving as their own instruments.

The identifying interpretation is the usual one for the demand system: instrumented variation moves the stock along the price-like dimension while the other included demand shifters define the conditional demand curve. Thus the price interpretation of α_{ft} does not come from treating arbitrary cross-sectional variation in firm size as a price shock; it comes from the instrumented component of market equity, conditional on the other characteristics in (13).

[19] Let V_{ft} denote the estimated variance–covariance matrix of the standardized coefficient vector and let indices 1 and k correspond to $(\alpha^{*,\mathrm{std}}, \beta^{BE,\mathrm{std}})$. Then $\mathrm{se}(\hat{\alpha}_{ft}) = \left(V_{11}/\sigma^2_{\log \mathrm{ME}} + V_{kk}/\sigma^2_{\log \mathrm{BE}} + 2V_{1k}/(\sigma_{\log \mathrm{ME}}\sigma_{\log \mathrm{BE}})\right)^{1/2}$. In practice, V_{1k} is often materially negative because $\log \mathrm{ME}$ and $\log \mathrm{BE}$ are highly correlated, so α can be more precisely estimated than α^{*} and β^{BE} individually.

[20] Position rows with no other fund left in the instrument after removing the fund's own term are dropped (1,793 rows in 790 fund-quarters).

Estimation criterion. Fund-quarter parameters are estimated by minimizing a quadratic form in the IV moment conditions:

$$\widehat{\theta}_{ft} \in \arg\min_{\theta} \left\| \sum_{j \in \mathcal{U}_{ft}} z_{fjt} \left(w_{fjt} - \mu_{fjt}(\theta) \right) \right\|^2, \tag{48}$$

where $z_{fjt} = (z^{\text{loo}}_{fjt}, \log \text{BE}_{jt}, X'_{jt})'$ stacks the leave-one-out price instrument with the exogenous characteristics of (44).

Standard errors. The standard errors linearize the level moment under the proportional benchmark. For one fund-quarter, let $u = w - \mu(\theta_0)$ collect the residuals over the fund's universe, $\xi_j = \varepsilon_j - 1$ the relative latent-demand departures, and X and Z the regressors and instruments, without a constant. Because the shares sum to one, to first order

$$u \approx \Omega\xi, \qquad \frac{\partial \mu}{\partial \theta'} = \Omega X, \qquad \Omega = \text{diag}(\mu) - \mu\mu', \qquad \text{Var}(\xi \mid X, Z) = \sigma^2 I. \tag{49}$$

Linearizing $Z'\{w - \mu(\hat{\theta})\} = 0$ around θ_0 gives $\hat{\theta} - \theta_0 \approx (Z'\Omega X)^{-1} Z'u$, and hence

$$\widehat{V}_{ft} = \hat{\sigma}^2_{ft} \, (Z'\Omega X)^{-1} \, (Z'\Omega^2 Z) \, (X'\Omega Z)^{-1}, \tag{50}$$

evaluated at the fitted shares. Because $\Omega z = \mu \circ (z - \sum_k \mu_k z_k)$, the instruments enter relative to their weighted mean under the fitted shares, which respects the adding-up constraint on the residuals, and each residual's variance is proportional to μ_j^2. The variance σ^2 is estimated from the fitted residuals $e = w - \mu(\hat{\theta})$ as

$$\hat{\sigma}^2_{ft} = \frac{e'e}{\text{tr}(\Omega^2) - \text{tr}\left[(Z'\Omega X)^{-1} Z'\Omega^3 Z \, (X'\Omega Z)^{-1} X'\Omega X \right]}, \tag{51}$$

where the second term in the denominator is the degrees-of-freedom correction for the estimated coefficients. $\widehat{V}_{ft}$ is the matrix V_{ft} of the delta-method calculation above.

These standard errors rest on the proportional benchmark. The reporting and implementation departures that motivate the level moment need not scale with the fitted position, and the variance formula does not accommodate them. It describes sampling uncertainty when every departure is latent demand proportional to the fitted position, not under the full error structure of (6).

A position-robust variance, which leaves the size of each residual unrestricted, is

$$\widehat{V}^{\text{rob}}_{ft} = (Z'\Omega X)^{-1} \Big(\sum_j e_j^2 \, \check{z}_j \check{z}_j' \Big) (X'\Omega Z)^{-1}, \qquad \check{z}_j = z_j - \sum_k \mu_k z_k, \tag{52}$$

which has the same expectation as (50) when $E[e_j^2] \propto \mu_j^2$.

Long-tail normalization. Score gradients scale with the fitted shares μ_j, which are tiny for tail positions, so the moments and their derivatives are numerically very small even though the estimand is well defined. To improve numerical scaling, I multiply the moment system by

a fund-specific scalar. Define

$$\mathrm{HHI}_{ft} \equiv \sum_{j \in \mathcal{U}_{ft}} w_{fjt}^2, \qquad n_{ft}^{\mathrm{eff}} \equiv \frac{1}{\mathrm{HHI}_{ft}}, \qquad s_{ft} \equiv n_{ft}^{\mathrm{eff}}. \tag{53}$$

Multiplying all weights by s_{ft} multiplies the moment system by a positive constant, which changes neither its root nor the minimizer of the quadratic criterion. It brings the moments to a scale at which the solver's tolerances are meaningful; it does not change the condition number of the Jacobian.

Sample selection. There are two stages of sample selection. Estimation first removes only mechanically undefined problems—insufficient dimensionality, failure of the estimating equation, or undefined sampling variance.[21] Over 2010Q2–2024Q4 I estimate 264,766 fund-quarters. No moment root is found for 10,695 of them (8,581 with no root, 249 with a singular Jacobian and 1,865 that fail the first-order check), and the sampling variance is undefined for a further 9,490. The 244,581 that remain have a usable price coefficient. These losses are 7.62% of fund-quarters and 3.94% of assets under management, concentrated in 2014–2015 among large income and balanced funds.

The empirical analysis then applies two ex ante quality screens: $F \geq 10$ for instrument relevance, following the conventional weak-instrument screen associated with Staiger and Stock (1997), and $n^{\mathrm{eff}} \geq 10$ for portfolio breadth. The statistic F is the squared t-ratio on the excluded instrument in the first-stage projection of the price regressor on the instrument and the remaining characteristics. It is a statement about instrument relevance, not a joint test that the demand slopes are zero, and I treat the threshold of 10 as a conventional screen rather than a critical value with formal justification for this nonlinear GMM problem. Two small groups also leave. Fund-quarters whose standard error on α_{ft} exceeds ten times the quarter's median are treated as estimation failures (272 fund-quarters). Fund-quarters that follow a price-weighted benchmark, flagged point in time from their holdings, have a different mechanical benchmark (339 fund-quarters). The resulting sample has 212,997 fund-quarters, which are 80.4% of those estimated and 89.9% of their assets under management.

For aggregate analyses, mandate coefficients are winsorized within quarter at the 1st and 99th percentiles.

C.4 The demand system with momentum

The baseline demand specification in Section 3.1 uses the characteristics of Koijen and Yogo (2019), which do not include momentum. The trading test in Section 4 nonetheless needs funds' tastes for momentum, because a stock's past return moves with its price during the trading quarter. I therefore re-estimate every fund-quarter with the stock's 12–2 month return added to X_{jt} as a seventh characteristic, winsorized and standardized within the quarter like the others, and check whether doing so changes the price coefficient. The sample period,

[21] Fund-quarters with no record in the CRSP fund file carry no assets and are excluded (893 estimated fund-quarters).

instrument, moment and screen are unchanged. The comparison uses the 212,683 screened fund-quarters with a usable estimate in both versions, out of 213,677.

Adding momentum leaves the price coefficients essentially unchanged. Across fund-quarters, the correlation between the two estimates of α_{ft} is 0.992, and it lies between 0.982 and 0.997 in every quarter. The holdings-weighted mean moves from 0.739 to 0.740. The median absolute change is 0.13 standard errors, and 2.6% of fund-quarters move by more than one standard error. Adding the characteristic costs essentially no precision: the median ratio of the two standard errors is 0.997. The index-fund benchmarks are also unchanged. The holdings-weighted mean coefficient of S&P 500 and total-market funds is 1.015 without momentum and 1.016 with it, and that of size- and sector-band index funds is 0.676 and 0.677.

The momentum version also yields each fund's estimated taste for momentum. Across funds, these tastes are negatively correlated with the price coefficient: -0.32 in the raw estimates and -0.43 after removing the covariance due to correlated estimation errors, about one-fifth of the raw covariance. This correlation across funds does not bias any one fund's price coefficient, which depends on how momentum co-varies with market equity within the fund's universe; the comparison above shows that dependence is small. It matters for the trading test, where a fund's response to momentum would otherwise load on its predicted price response. Section 4 uses these tastes, shrunk toward the fund's own estimates of the three preceding quarters in the same way as $\hat{\alpha}_f$, to build the momentum control in Table 3. Because the two versions give nearly identical price coefficients, the predicted trade uses the baseline estimates.

C.5 Stocks sorted on non-passive elasticity

Table 8 repeats Table 1 after excluding passive holders. Section 3.2.2 discusses the comparison.

Table 8: **Stocks by quintile of non-passive elasticity.**

	$\bar{\kappa}_{NP}$ quintile				
	Low	2	3	4	High
Elasticity					
$\bar{\kappa}_{NP}$	0.373	0.513	0.600	0.673	0.774
$\bar{\kappa}_{NP}$ four quarters earlier (same stocks)	0.406	0.532	0.608	0.672	0.747
$\bar{\kappa}_{NP}$, 2010–2014	0.420	0.557	0.637	0.701	0.791
$\bar{\kappa}_{NP}$, 2015–2019	0.370	0.511	0.604	0.679	0.776
$\bar{\kappa}_{NP}$, 2020–2024	0.331	0.473	0.560	0.639	0.755
κ_j	0.270	0.398	0.481	0.546	0.628
Elasticity of the stock's holders, by group					
Growth funds	0.338	0.518	0.624	0.702	0.799
Value funds	0.339	0.417	0.455	0.492	0.509
Other active funds	0.419	0.546	0.626	0.682	0.736
Group mix only (shares $\times$ market-wide group elasticity)	0.424	0.425	0.426	0.427	0.425
Holder base					
Mutual-fund ownership (% of shares)	29.75	34.06	37.26	39.78	39.59
Number of screened fund holders	553	368	300	278	248
Passive funds (% of mutual-fund $)	38.13	36.73	35.67	33.37	30.15
Growth funds (%)	17.62	16.98	17.41	21.61	32.57
Value funds (%)	12.55	15.61	15.24	12.66	7.43
Other active funds (%)	31.70	30.69	31.68	32.36	29.85
Firm					
S&P 500 members (%)	64.65	34.19	13.39	4.67	1.15
NYSE size decile	8.06	6.39	5.46	5.12	4.76
Market cap ($bn)	53.9	8.5	4.1	3.1	2.4
BE/ME	0.508	0.602	0.584	0.500	0.372
Momentum decile	5.78	5.75	5.77	5.89	6.22
Profitability	0.310	0.266	0.273	0.269	0.256
Investment	0.139	0.137	0.130	0.151	0.214
Market beta	1.16	1.25	1.26	1.24	1.26
Dividend yield (%)	1.76	1.60	1.45	1.15	0.73
Return volatility (annualized %)	36.77	38.74	38.06	37.62	39.71
Stock-quarters	22,789	22,756	22,751	22,756	22,777

Notes. The sample is that of Table 1. Stocks are sorted within each quarter on $\bar{\kappa}_{NP}$, which renormalizes the holder weights of Growth, Value and other active funds within each stock. These groups exclude conventional index funds but keep style, dividend and factor index funds. It is undefined for five of 113,834 stock-quarters, which have no such holders. Group elasticity rows cover stocks with holders in that group. The group-mix row prices each stock's shares of the three groups at the quarter's market-wide dollar-weighted group elasticities. Each cell is the time-series average of quarterly equal-weighted means. Fund groups use the point-in-time Lipper classification, and fund shares are shares of the dollars held by screened mutual funds.

D Sampling Uncertainty in the Elasticity Trend

The shaded areas in Figure 5 are 95% bands of ± 1.96 standard errors for the sampling uncertainty in the fund coefficients. Standard errors cluster the funds' influence by stock, so they allow errors to be correlated across funds that hold the same stocks, and each fund's own variance is set to its fixed-design variance. They assume independence across stocks and do not cover misspecification. A few very large stocks dominate the variance, with about eight effective stock clusters in the median quarter for all funds, so the width of a band in any one quarter is itself imprecise. Under these assumptions, the fall in the dollar-weighted elasticity from 2010Q2 to 2024Q4, 0.23, has a standard error of 0.015, and its linear trend of -0.014 per year has a standard error of 0.0013. For non-passive funds, the fall of 0.20 has a standard error of 0.022 and the trend of -0.013 per year a standard error of 0.0019.

Construction. The dollar-weighted elasticity in quarter t is $\bar{\kappa}_t = \sum_f \sum_j A_{fjt}\kappa_{fjt}/A_t$, where A_{fjt} is fund f's holding of stock j in dollars and A_t the total. Because $\kappa_{fjt} = 1 - \alpha_{ft}(1 - w_{fjt})$, its derivative with respect to the fund's price coefficient is $-\lambda_{ft}$, with $\lambda_{ft} = \sum_j A_{fjt}(1 - w_{fjt})/A_t$. From the linearization in Appendix C.3, each fund's estimate has the influence representation $\hat{\alpha}_{ft} - \alpha_{ft} \approx \sum_{j \in \mathcal{U}_{ft}} h_{fjt}$, with $h_{fjt} = c'(Z'\Omega X)^{-1}\check{z}_{fjt}\,(w_{fjt} - \mu_{fjt})$, where c maps the coefficient vector into the price coefficient α_{ft} and $\check{z}_{fjt} = z_{fjt} - \sum_k \mu_{fkt} z_{fkt}$. Each fund's influences are rescaled so that $\sum_j h_{fjt}^2$ equals the fund's fixed-design variance of $\hat{\alpha}_{ft}$ from (50). The influences are then summed across funds within each stock, $S_{jt} = \sum_f \lambda_{ft} h_{fjt}$, and the variance of $\bar{\kappa}_t$ is estimated as $\sum_j S_{jt}^2$. Estimation errors are thereby allowed to be correlated across funds through the stocks they hold, while different stocks are treated as independent. The sums run over every stock in a fund's universe, including those it does not hold.

The endpoint change and the trend are linear combinations $\sum_t c_t \bar{\kappa}_t$ of the quarterly series, with $c_t = \pm 1$ at 2024Q4 and 2010Q2 for the change and $c_t = (t - \bar{t})/\sum_s (s - \bar{s})^2$, with time in years, for the trend. Their variance is estimated as $\sum_j (\sum_t c_t S_{jt})^2$, which clusters by stock across quarters: errors attached to the same stock may be correlated over time. The non-passive series is built in the same way from the influences of non-passive funds, with weights relative to non-passive dollars.

E Robustness of the Out-of-Sample Test

Table 9 re-estimates the specification of Table 3, panel C, column (4), changing one element at a time: the construction of unexpected flow, the funds whose flow-induced trading is left out of the instrument, and the shrinkage of the fund's estimated tastes. It also re-estimates the specification in subperiods and by stock size.

Table 9: **Robustness of the trading response.**

	γ	s.e.	First-stage F	Observations (millions)
Table 3, Panel C, column (4)	0.682	(0.052)	90.9	20.9
A. Construction of unexpected flow				
Forecast from own flow and return lags	0.655	(0.063)	72.0	20.9
Adding quarter effects and class-return lags	0.615	(0.047)	93.6	20.9
Winsorized returns in the forecast	0.651	(0.062)	69.5	20.9
Rolling 12-quarter forecast window	0.620	(0.063)	64.3	20.9
Scaled by Lou's price-sensitivity factors	0.660	(0.063)	68.8	20.9
Lou's denominator	0.660	(0.054)	86.6	20.9
Raw flow for funds without forecast lags	0.686	(0.064)	68.9	20.9
B. Funds left out of flow-induced trading				
None	0.659	(0.050)	98.7	20.9
Own company only	0.658	(0.051)	99.4	20.9
Own Lipper class only	0.673	(0.051)	99.4	20.9
C. Fund tastes				
Unshrunk estimates	0.610	(0.049)	88.5	20.9
D. Subperiods				
2012Q3–2016Q4	0.876	(0.108)	54.2	6.0
2017Q1–2024Q4	0.619	(0.056)	52.8	14.9
Excluding 2020Q2	0.693	(0.057)	89.3	20.5
Excluding 2020Q2–2021Q1	0.718	(0.070)	61.7	19.0
E. Stock size at $t-1$				
Largest 200 stocks	0.749	(0.083)	65.5	5.2
Stocks ranked 201–1,000	0.742	(0.070)	88.1	10.0
Smaller stocks	0.651	(0.118)	63.8	5.7

Notes. Each row re-estimates the specification of Table 3, Panel C, column (4): the trade Trade_{fjt} on the predicted trade $\widehat{\text{Trade}}_{fjt} = (\hat{\alpha}_f - 1)(\Delta p_{jt} - \Delta \bar{p}_{ft})$, instrumented with $(\hat{\alpha}_f - 1)$ times unexpected-flow FIT, with Lipper class $\times$ price change, own-company FIT, the stock characteristics (including the momentum and book-to-market terms) and fund-quarter and stock-quarter fixed effects, weighted by drift weights. Panels A–C change one element of the price instrument or of the fund's estimated tastes; the momentum instrument is unchanged, and the book-to-market term is instrumented with the same FIT as the predicted trade, with the corresponding own-company FIT as a control. In panel A, the main forecast of unexpected flow follows Lou (2012); the alternatives change the forecast, the scaling or the flows used. Panel B changes the funds whose flow-induced trading is left out (main: own company and Lipper class). In panel C, the fund's price sensitivity $\hat{\alpha}_f$, momentum taste and book-to-market coefficient are the raw quarter-$(t-2)$ estimates instead of estimates shrunk toward the fund's own estimates of the three preceding quarters. Panels D and E re-estimate the same specification on subsamples, with their own fixed effects; size groups rank stocks by market capitalization at $t-1$. Standard errors are clustered by fund, stock and quarter; the first-stage F is the effective F under the same clustering.