Risk Versus Transaction Costs: Understanding the Cross-Section of Institutional Flows*

Carter Davis† Mahyar Kargar‡ Jiacui Li§

February 11, 2026

Abstract

While the asset pricing literature has devoted enormous attention to the cross-section of returns, far less is known about the cross-section of flows—the systematic way in which portfolio reallocation differs across stocks. When institutions perceive a change in expected returns, they adjust their portfolios, but the magnitude of adjustment is constrained by either risk aversion—as in classical models—or by trading frictions. Which force dominates has first-order implications for how we model institutional demand. We document that institutional portfolio weight adjustments scale monotonically with stock market capitalization, a pattern that is inconsistent with classical models. Incorporating transaction costs into the portfolio choice framework resolves this tension: for realistically-sized mutual funds and 13F institutions, transaction costs dominate risk aversion in limiting position sizes, and because larger stocks are more liquid, institutions optimally tilt their trading toward them. Our model quantitatively matches the cross-sectional scaling of flows with market capitalization and generates additional predictions—about how this scaling varies with fund size and investment universe liquidity—that bear out in the data. Our results show that trading frictions, not risk, are the dominant force shaping the cross-section of institutional flows.

Keywords: Cross-section of flows, Portfolio choice, Transaction costs, Demand system asset pricing
JEL Classification: G11, G12, G14.

*We thank Robert Richmond, Zhaoxin Shi, and Dejanir Silva for comments.
†The Ohio State University Fisher College of Business; carter.davis@fisher.osu.edu
‡Gies College of Business, University of Illinois Urbana-Champaign; kargar@illinois.edu
§David Eccles School of Business, University of Utah; jiacui.li@eccles.utah.edu

1 Introduction

A central but underexplored feature of financial markets is the cross-section of flows: the systematic way in which trading activity and portfolio reallocation differ across stocks. Asset pricing research has focused far more on the cross-section of returns—why some securities earn higher average returns than others. By contrast, how trading and rebalancing vary across assets has received much less attention. Flows are nonetheless central to equilibrium, revealing where investors allocate capital, how aggressively they rebalance, and which assets absorb demand and supply shocks. Because information, risk-sharing, and liquidity demand are reflected in prices largely through flows, understanding their cross-sectional structure is crucial.

What drives that structure remains an open question. When an institution's view of expected returns changes, it adjusts its portfolio, but the size of that adjustment is constrained by risk aversion—as in classical portfolio choice—and by the cost of trading. Whether risk or frictions are the binding constraint is an empirical question. Characterizing when transaction costs dominate risk considerations, or vice versa, provides a new framework for organizing empirical evidence on trading behavior and sharpens our understanding of the economic forces shaping real-world portfolio choice.

In this paper, we show that, for large institutions, trading frictions are the dominant force shaping the cross-section of institutional flows. We document a robust stylized fact: for active mutual funds and 13F institutions, portfolio weight adjustments scale monotonically with stock market capitalization—the larger the stock, the more institutions trade it. In contrast, classical portfolio choice models predict that active adjustments should be approximately flat in stock size. We resolve this puzzle by incorporating transaction costs into the portfolio choice framework. For realistically-sized institutions, transaction costs dominate risk aversion in limiting position sizes; because larger stocks are more liquid and less costly to trade, the model predicts that institutions

tilt their trading toward large stocks—matching the data. The framework also generates testable predictions about how this scaling varies with fund size and investment universe liquidity, which we confirm empirically.

We begin by documenting the stylized fact in detail. Using quarterly holdings of active U.S. domestic equity mutual funds and 13F institutions from 1990 to 2023, we measure discretionary portfolio weight adjustments net of flow-induced trading. Within each fund-quarter, we sort stocks into bins by lagged market capitalization. The average magnitude of discretionary trades rises monotonically with stock size: stocks in the smallest bin are traded at roughly 0.2% of portfolio weight per quarter, while stocks in the largest bin are traded at roughly 0.6%. The relationship holds for alternative measures of active positioning—including deviations from benchmark weights—and extends to the broader universe of 13F institutions. Critically, controlling for return volatility does not attenuate the pattern.

Classical portfolio choice models struggle to explain these patterns. Under static mean-variance optimization, the optimal portfolio weight on stock i is approximately $w_{i,t}^{\text{MV}} \approx \mu_{i,t}/(\gamma\sigma_{i,t}^2)$, where $\mu_{i,t}$ and $\sigma_{i,t}$ are subjective expected return and volatility and γ is risk aversion.[1] With volatility relatively stable over time and subjective expected returns displaying limited cross-sectional dispersion (Davis et al., 2025; Hou, Xue, and Zhang, 2020), trading volume in portfolio weight terms is driven primarily by changes in perceived expected returns: $\Delta w_{i,t}^{\text{MV}} \approx \Delta\mu_{i,t}/(\gamma\sigma_i^2)$. This yields a counterfactual prediction. Larger stocks exhibit less return predictability—both in statistical terms and in analyst forecast dispersion. To explain why institutions trade large stocks more, the MV framework would require assuming much greater expected return variation in large stocks, which runs counter to decades of evidence (e.g., Hou et al., 2020).

[1] This expression ignores cross-stock substitutions which, quantitatively, is small when studying how trading volume scales across stocks. Appendix A in Davis, Kargar, and Li (2025) shows that the correct expression is $w_{i,t}^{\text{MV}} = \frac{\mu_{i,\text{unspanned},t}}{\gamma\sigma_{i,\text{unspanned},t}^2}$, where we define the "unspanned" part of stock i return as the residual after a linear projection onto all the other stocks.

The gap between the frictionless MV model and the data is quantitatively large. Calibrating risk aversion γ to match trading on the smallest decile of stocks leads the model to underestimate trading on the largest decile by over 30-fold. Common extensions—benchmarking (Basak and Pavlova, 2013), intertemporal hedging (Merton, 1971; Campbell and Viceira, 1999)—do not resolve this shortcoming in any quantitatively meaningful way. We show that dynamic hedging considerations, while theoretically relevant, are quantitatively small for the cross-sectional pattern of trading volume. The frictionless MV framework, in its canonical form and leading variants, cannot match the data.

The missing ingredient is the internalization of transaction costs (TC). Standard asset pricing models assume investors are price takers. For retail investors, this assumption is reasonable; for large institutional investors whose trades move prices, it is not. We modify the MV portfolio choice framework by incorporating an explicit minimization over the price impact of trade. The key insight is that transaction costs scale inversely with stock size: larger stocks are more liquid and less costly to trade. This introduces a natural tilt toward large stocks in optimal portfolio adjustments—the same friction that discourages trading in small stocks encourages it in large ones. For realistically-sized mutual funds and 13F institutions, transaction costs dominate risk aversion in limiting position sizes. When calibrated using standard price impact estimates from the financial industry, the model quantitatively matches the cross-sectional scaling of flows with market cap.

The framework generates additional testable predictions. The importance of transaction costs depends on institution size and investment universe liquidity. Larger institutions, or those investing in less liquid stocks, should exhibit a stronger relationship between trade magnitude and stock size. Small investors for whom price impact is negligible should not. We test these predictions and find strong support: institutions with larger assets under management, or those holding less liquid portfolios, display a steeper scaling of trading with market capitalization. The opposite holds for

retail investors (Barber and Odean, 2000), whose trading does not exhibit a clear relationship with stock size—consistent with transaction costs being negligible at that scale.

Taken together, these results provide a clear answer to whether risk or frictions dominate the cross-section of flows: for institutional investors, transaction costs are the primary force. Risk aversion plays a role—it governs flows in the smallest and least liquid stocks where price impact is modest—but for the vast majority of institutional trading activity, it is the friction of moving prices, not the discomfort of bearing risk, that shapes where capital is allocated. This finding reframes the standard view in portfolio theory, which treats transaction costs as a second-order correction to risk-based portfolio choice.

The cross-section of flows has received less systematic attention than the cross-section of returns. Our paper contributes to several strands that touch on it. The first concerns the foundations of portfolio choice. The mean-variance framework of Markowitz (1952) and its intertemporal extensions (Merton, 1971, 1973; Campbell and Viceira, 1999; Campbell, Chan, and Viceira, 2003) have long served as the workhorse for theoretical asset pricing. These models abstract from transaction costs and assume investors are price takers. Constantinides (1986) showed that transaction costs can have first-order effects on portfolio choice in a partial equilibrium setting. More recently, Gârleanu and Pedersen (2013) and Gârleanu and Pedersen (2016) have developed dynamic portfolio choice models with predictable returns and transaction costs, emphasizing the trade-off between trading intensity and execution costs. We differ by focusing on the cross-sectional implications: when and where do transaction costs versus risk govern how investors allocate trades across stocks? Our key contribution is to show that for large institutions, transaction costs dominate—and to characterize the institutional and market conditions under which this dominance holds.

A second related literature examines institutional investor behavior and its implications for

asset prices. Keim and Madhavan (1995) document that institutions spread trades over time to reduce execution costs. Edelen, Evans, and Kadlec (2013) argue that transaction costs, including price impact, form a larger drag on active fund performance than expense ratios. Pollet and Wilson (2008) and Cremers and Petajisto (2009) study how fund size affects portfolio composition and activity. Basak and Pavlova (2013) analyze equilibrium asset prices when institutions care about performance relative to benchmarks. Our paper complements this work by providing a structural framework that links institutional size, transaction costs, and the cross-sectional pattern of trading—and that predicts when each force dominates.

A third literature has studied flows and demand. Lou (2012) examines how fund flows affect return predictability. Koijen and Yogo (2019) estimate institutional demand as a function of stock characteristics and use the estimated demand system to study equilibrium implications. Gabaix and Koijen (2022) propose an inelastic markets hypothesis in which demand elasticities play a central role. Davis (2025) contrast demand specifications used in this literature with classical portfolio choice. Our paper takes a different angle: rather than specifying demand functions directly, we ask what economic forces—risk versus transaction costs—determine the cross-section of flows: the systematic way in which trading activity and portfolio reallocation differ across stocks. We show that the answer varies with investor size and market liquidity.

Finally, our paper relates to the market microstructure literature on price impact. Almgren, Thum, Hauptmann, and Li (2005) and Frazzini, Israel, and Moskowitz (2018) provide empirical estimates of equity market impact. Industry practice, as reflected in Grinold and Kahn (2000), integrates transaction cost modeling into portfolio construction. Novy-Marx and Velikov (2016) and Novy-Marx and Velikov (2019) study how trading costs vary across anomalies and how to mitigate them. We take these estimates as inputs and show that incorporating them into a standard

portfolio choice framework yields testable predictions about the cross-section of flows—predictions that distinguish when transaction costs versus risk govern investor behavior.

The remainder of the paper proceeds as follows. Section 2 documents the stylized facts and establishes the failure of canonical portfolio choice models. Section 3 presents the TC-augmented model and its testable implications. Section 4 tests these implications using institutional holding and trading data. Section 5 concludes.

2 Comparing Canonical Portfolio Choice Models to Data

This section starts by documenting stylized facts about how institutional investor trading volume scales across stocks with different sizes. We then show that canonical models of portfolio choice make counterfactual predictions about the trading behavior of institutions.

2.1 Stylized facts about discretionary institutional trading

Understanding the elasticity of markets requires knowing how actively investors trade. Towards this end, we first take a preliminary look at the sizes of discretionary portfolio adjustments of active U.S. domestic equity mutual funds from 1990 to 2023.[2]

To compute discretionary portfolio weight adjustments, we compute:

$$\Delta w_{i,j,t} = \frac{(s_{i,j,t} - s_{i,j,t-1}) \cdot p_{i,t-1}}{\text{AUM}_{j,t-1}} - \underbrace{f_{j,t} \cdot w_{i,j,t-1}}_{\text{flow-induced trading}} \tag{1}$$

[2]We obtain quarterly holdings from 13F filings and merge with CRSP to obtain stock characteristics. We use the CRSP mutual fund database, linked via MFLINKS of Wermers (2000), to focus on domestic equity funds and filter out index funds and ETFs. To ensure sufficient variation in holdings within each fund, we examine funds with at least 50 stock holdings and 10 million USD in assets under management.

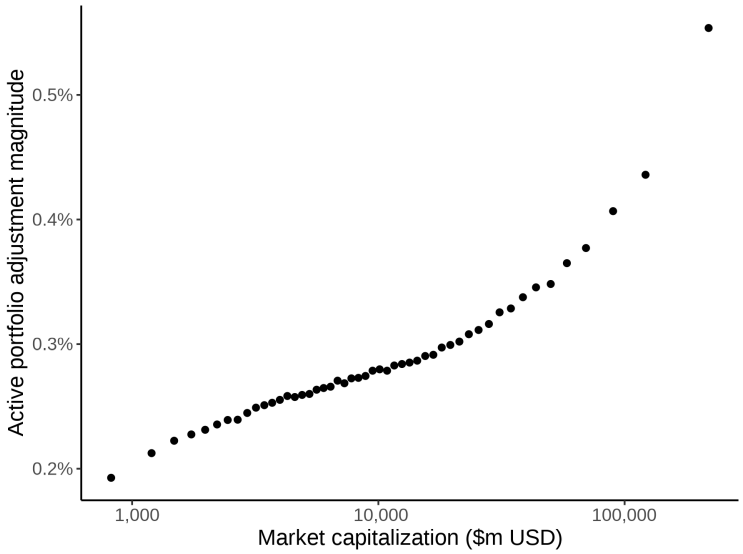


Figure 1. How discretionary mutual fund trade magnitudes scale with stock size.
For active U.S. domestic equity mutual funds from 1990 to 2023, within each fund-quarter, we sort stock holdings into 50 bins by lagged stock market capitalization. We plot the absolute value of active quarterly portfolio adjustments ($|\Delta w_{i,j,t}|$), defined as the change in portfolio weight due to discretionary trading (equation (1)), by bin.

where $s_{i,j,t}$ is the number of split-adjusted shares held by fund j in stock i at the end of quarter t, $p_{i,t-1}$ is the lagged stock price, $\text{AUM}_{j,t-1}$ is the lagged fund AUM, $f_{j,t}$ is the fund flow in quarter t as a fraction of lagged AUM, and $w_{i,j,t-1}$ is the lagged portfolio weight in stock i. Following Sammon and Shim (2023), we subtract the flow-induced trading term to isolate discretionry trades that are not driven by funds mechanically adjusting positions to accommodate flows.[3] Therefore, $\Delta w_{i,j,t}$ serves as a measure of the discretionary trading decisions made by fund managers.

Within each fund-quarter, we sort stocks into 50 bins based on the lagged stock market capitalization. Figure 1 plots the average magnitude of discretionary trades by bin, and it shows that mutual funds conduct larger discretionary trading in larger stocks. Stocks in the bottom bin,

[3] Lou (2012) shows that funds approximately scale pre-existing positions in response to flows. In unreported robustness checks, we verify that using different flow-to-trading coefficients for inflows versus outflows does not materially impact our conclusions.

which on average have less than 1 billion dollars of market capitalization, have a mean trade size of slightly below 0.2%, while stocks in the top bin are traded around 0.6% per quarter.

This finding—that institutions trade more in large stocks—is robust over time and across subsamples. In unreported robustness checks, we verify that this pattern holds for mutual funds of various sizes and styles across all years. We obtain similar results when using alternative measures of active position-taking, such as portfolio deviations from benchmarks, and find that the pattern also holds for the broader universe of 13F institutions. The next section shows that this pattern is difficult to reconcile with classical portfolio choice models.

2.2 Testing canonical models: baseline

In this section, we derive the prediction of static mean-variance (MV) portfolio choice and argue that it cannot explain the patterns in Section 2.1.

MV model prediction. Consider an investor choosing portfolio weight vector $w_t = (w_{1,t}, ..., w_{N,t})'$ over N stocks. The risk-free rate is set exogenously, and risky assets have subjective expected return vector μ_t and covariance matrix Σ_t. The investor has Mv utility over her performance relative to a benchmark portfolio w_b:

$$U = (w_t - w_b)'\mu_t - \frac{\gamma}{2}(w_t - w_b)'\Sigma_t(w_t - w_b)$$

where $\gamma > 0$ is the risk aversion parameter. Note that this nests the classical MV model as a special case of $w_b = 0$.

The optimal portfolio choice is given by the familiar expression of $w_t = w_b + \frac{1}{\gamma}\Sigma_t^{-1}\mu_t$. To study how portfolio weights vary across stocks, for each stock i, this can be expressed using unspanned

expected returns and variance (Stevens, 1998; Davis et al., 2025):

$$w_{i,t} = w_{i,b} + \frac{\mu_{i,\text{unspanned},t}}{\gamma \sigma^2_{i,\text{unspanned},t}},$$

$$\text{with} \begin{cases} \mu_{i,\text{unspanned},t} & \equiv \mu_{i,t} - \underbrace{\beta_{i,-i,t} \cdot \mu_{-i,t}}_{\text{spanned return}}, \\ \sigma^2_{i,\text{unspanned},t} & \equiv \sigma^2_{i,t} - \underbrace{\beta'_{i,-i,t} \Sigma_{-i,-i,t} \beta_{i,-i,t}}_{\text{spanned variance}}, \end{cases}$$

where $\beta_{i,-i,t}$ is an $(N-1) \times 1$ vector of coefficients obtained from linearly projecting $r_{i,t}$ onto the vector of all stocks except i, which we denote as $r_{-i,t}$. Similarly, the notation of $\mu_{-i,t}$ refers to the μ_t vector without stock i, and $\Sigma_{-i,-i,t}$ is the $(N-1) \times (N-1)$ sub-covariance matrix with i purged from rows and columns. Intuitively, $\beta'_{i,-i,t} r_{-i,t}$ is the return of the best linear replicating portfolio of stock i, and the part of stock i expected returns (variances) not captured by this replicating portfolio is called "unspanned". The "spanned" components effectively capture the cross-stock substitution in MV portfolio choice.

In unreported analysis, we show that $\sigma_{i,\text{unspanned},t}$ is approximately equal to the notion of "idiosyncratic volatility" after projecting each stock onto commonly used factors that explain covariances. In the U.S. stock market, most of stock-level return volatility is unspanned, and the unspanned volatility is rather stable over time and thus cannot drive large portfolio shifts on a quarter-to-quarter basis. Therefore, most portfolio shifts must be associated with changes in (possibly subjective) expectation of unspanned returns. Further, to focus on how trade sizes scale across stocks within institution-period, we can omit the risk aversion term:

$$\Delta w_{i,t} \propto \frac{\Delta \mu_{i,\text{unspanned},t}}{\sigma^2_{i,\text{unspanned}}} \tag{2}$$

In other words, institutions tend to trade more in stocks in which their (subjective) expected returns have changed more, with a scaling factor based on unspanned variance.

This prediction, while intuitive, have trouble matching the stylized facts in Section 2.1. While volatilities play a role in (2), Section 3 shows that controlling for volatilities have little impact on how trade magnitudes scale with stock sizes. Therefore, within the MV framework, to explain the stylized fact, we have to assume that institutions have much more volatile expected returns on large stocks.

However, this goes against common intuition and decades of research. It is well documented that return predictability tends to be weaker, not stronger, in larger stocks (Hou et al., 2020), a fact that we corroborate in Appendix A.2. To investigate the variation of *subjective* expected returns, Appendix A.2 also extract implied return forecasts from analyst price targets, and the subjective forecasts are also less volatile in large stocks. While we agree that, due to having different information and expertise, mutual fund managers can have rather heterogeneous beliefs about expected returns. Having said that, it would be *rather behavioral* for them to believe that large stocks contain much larger and more time-varying alpha opportunities. Therefore, we argue that traditional MV portfolio choice is not consistent with the stylized facts in Section 2.1.

To quantitatively assess the degree of mistmatch between model and data, we provide a back-of-the-envelope calculation. Section 4.2 shows that $|\Delta w_i|$ scales roughly proportionally to $M_i^{\beta_w}$ with $\beta_w \approx 0.3$ across stocks in mutual funds trades, after controlling for the impact of volatilities. Appendix A.2 shows that $|\Delta \mu_i| \sim M_i^{-\beta_\mu}$ with $\beta_\mu \approx 0.2$. This means that the MV model is off by a factor of $M_i^{\beta_w+\beta_\mu} \approx M_i^{0.5}$. When comparing bottom to top stock size deciles, the market capitalization differs by over 1,000 times. Therefore, if the risk aversion γ is calibrated by making the MV model match the trading magnitudes on the bottom decile, it would underestimate trading volume in the top decile by a factor of $1{,}000^{0.5} \approx 31$.

We clarify two points about the puzzle we document. First, we focus on how portfolio *changes* scale with stock sizes, rather than portfolio *levels*, because one can easily rationalize the fact that funds hold large stocks more by noting that they may follow benchmarks that lean towards large cap stocks. This can be easily captured within the MV framework by setting w_b to be related to market cap. However, benchmarks are differenced out when examining trades. Second, while our finding is also true for other institutions, we start by focusing on the discretionary trades of active mutual funds to guard against the concern that trades may be driven by flows in or out of market cap-based indices.

Does this discrepancy between classical models and realistic portfolio choice matter? We argue that it does, as the two different scaling behaviors predict qualitatively different patterns about demand elasticities across stocks. As pointed out in Davis (2025), in equilibrium models with MV portfolio choice, investors exhibit significantly lower demand elasticities in large stocks, which predicts that "price multipliers"—defined as the percentage price change for 1% of information-less demand of shares—are significantly higher in large stocks. This is counterfactual, as empirically estimated price multipliers are not significantly different across stocks—which is broadly consistent with the empirically observed portfolio behavior.

2.3 Testing canonical models: extensions

We argue that the core puzzle in Section 2.2 cannot be addressed by frictionless extensions of portfolio choice models.

Dynamic extensions Since Merton (1971) and Campbell and Viceira (1999), researchers have considered how investors choose portfolios in dynamic settings. In such settings, in addition to the

static mean-variance portfolio, investors also have a hedging portfolio. In vector form, it means:

$$\Delta w_t \approx \underbrace{\frac{1}{\gamma}\Sigma^{-1}\Delta\mu_t}_{\text{static}} + \underbrace{\Delta w_t^{\text{hedging}}}_{\text{hedging}}$$

Our analysis thus far only has the static term and does not consider the hedging term. We argue that, under realistic parameters, considering the hedging term does not materially alter our cconclusions.

Davis et al. (2025) shows that, when examining demand patterns for individual stocks, the hedging term, which arises from the need to hedge aggregate investment opportunities or consumption, tend to be quantitatively small. In Appendix A.4, we further examine the impact of hedging by making use of the insight in Lo and Wang (2006). They show that an institution's trades will exhibit a "two-fund separation" structure: hedging-related trades should be aligned along a "hedging portfolio" that is optimized for the fund at hand, although the exact portfolio will differ from fund to fund. We use principal component analysis on each fund's trades to extract factor structures. We observe that mutual funds' trades do not have strong factor structures; most trades appear to be stock-specific. After removing those PC factors, the puzzle remains: funds tend to trade more in stocks that are larger and more liquid.

Models with infrequent belief updating. In models where fund managers do not update their beliefs about all stocks, they will only trade those that they have updated their beliefs about. If fund managers update beliefs about small stocks infrequently, then they will only trade small stocks occasionally, and this may explain our findings in Section 2.1.

Under this explanation, the low trading sizes in small stocks is a *compositional* effect: small

stocks often have zero trades. To guard against this possibility, we note that our conclusions in Section 2.1 are robust to restricting attention to stocks with a non-zero amount of trading.

3 Internalization of Transaction Costs

In this section, we propose that the stylized facts in Section 2 are best explained by institutions internalizing their transaction costs (TC)—broadly construed to include both fixed costs of transactions (commission, and bid-ask spreads) and price impacts. This naturally leads institutions to trade more in large stocks due to the higher liquidity.

We start by noting that while many equilibrium asset pricing models assume infinitesimal investors who are competitive, there is ample evidence that TCs form a significant drag on institutional returns, and that the institutions try to reduce TCs. Edelen et al. (2013) argue that, when accounting for price impact, TCs form a larger drag on the performance of active mutual funds than expense ratios. Keim and Madhavan (1995) show that institutions spread out their trades over time to reduce TCs. Institutional investors leverage pre-trade analytics from specialized data providers, such as ITG (now Virtu), to calibrate their portfolio selections against expected execution costs. Practitioner-focused literature offers explicit approaches for integrating TC avoidance into the portfolio construction process (e.g. Grinold and Kahn, 2000), a focus also reflected in recent academic research (Gârleanu and Pedersen, 2013; Novy-Marx and Velikov, 2016, 2019).

3.1 Model set up

We now extend the MV portfolio choice framework to incorporate TC internalization. Under realistic parameter assumptions, the model predicts that investors will tilt their trading activity toward larger, more liquid stocks. Furthermore, this relationship between trade magnitude and

stock size is stronger for institutional investors managing larger AUMs or operating in less liquid stock universes. We begin by presenting a generic model before deriving testable implications based on a simplified special case.

Consider an institution with assets under management A with N risky assets in its investment universe. Let μ_t be the vector of subjective expected returns and Σ be the return covariance matrix. Let w_b, often a value-weighted portfolio, be the portfolio that the institution is benchmarked to. We model the institution as conducting a TC-augmented MV optimization:

$$\max_{w_t} (w_t - w_b)'\mu_t - \frac{\gamma}{2}(w_t - w_b)'\Sigma(w_t - w_b) - \underbrace{\frac{1}{2}(w_t - w_{t-1})'\Lambda(w_t - w_{t-1})}_{\text{transaction cost}} \tag{3}$$

where the last term captures TC, following the formulation in Gârleanu and Pedersen (2013): Λ is an $N \times N$ matrix and w_{t-1} denotes the vector of portfolio weights in the previous period. The vector of transaction costs (in return terms) is modeled as $\frac{1}{2}\Lambda(w_t - w_{t-1})$, and we further pre-multiply it by the vector of trades $(w_t - w_{t-1})'$ to arrive at the total cost.

What should we assume about Λ which governs price impacts? Because cross-stock price impact is small, as a baseline, we assume that Λ is a diagonal matrix. Industry-oriented research on TC usually assume away cross-stock price impact (e.g. Almgren et al., 2005; Frazzini et al., 2018). The fact that cross-stock price impact is small is also confirmed in academic work (Chaudhry and Davis, 2026).

As for the diagonal terms λ_i, as a baseline, we assume:

$$\lambda_i = \frac{\lambda_0 A^{\alpha_A}}{M_i^{\alpha_M}}, \tag{4}$$

where λ_0 is a constant, M_i is the market capitalization, and A is the AUM of the fund. Appendix A.1

provides a derivation for this specification and empirically estimates that $\alpha_A \approx 0.58$ and $\alpha_M \approx 0.79$ in U.S. stocks. The direction of scaling matters for our qualitative derivations. Importantly, price impact is larger if the fund is larger ($\alpha_A > 0$), because 1% of portfolio reallocation for a larger fund means more trading in dollar terms. At the same time, price impact is smaller if the underlying stock is larger ($\alpha_M > 0$), as larger stocks have more liquidity.

This formulation nests as a special case where transaction costs entirely arise from price impacts—which would be appropriate for large institutions—and where stock have equal "price multipliers": $\alpha_A = \alpha_M = 1$. To see this, recall that price multipliers are defined as the percentage price impact for each demand shock of 1% of total market capitalization. Suppose stocks share the same multiplier λ_0, then we simply need to set $\lambda_0 = 2\lambda_0$ to capture this case.[4]

3.2 Testable predictions

The solution to (3) is given by a modification to the familiar expression for efficient portfolios:

$$w_t = w_b + \underbrace{(\gamma\Sigma + \Lambda)^{-1}}_{\text{risk aversion + TC}} (\mu_t + \Lambda w_{t-1}).$$

To gain intuition, we consider a special case where $w_{t-1} = w_b$, that is, the initial portfolio position is the same as the benchmark. We further assume that the covariance has a single factor structure $\Sigma = bb'\sigma_f^2 + D$ where D is diagonal, following Sharpe (1963). Under these assumptions, Appendix B.1 shows that the optimal portfolio adjustment in stock i can be written as

$$w_{i,t} - w_{i,t-1} = \frac{\mu_{i,\text{unspanned},t}}{\gamma\sigma^2_{i,\text{unspanned},t} + \lambda_i}.$$

[4]Note that the dollar value of trading by the institution in stock i is $A \cdot \Delta w_i$. Thus, the total price impact should be $\frac{\lambda_0 A \Delta w_i}{M_i}$.

Relative to the frictionless MV case, the only difference is the presence of λ_i—the TC parameter—in the denominator. Intuitively, we can think of λ_i as inflating the unspanned variance of stock i, and thus, stocks with higher TC parameters will be traded less.

As shown in (4), the magnitude of price impacts parameters is increasing in the size of the institution A and decreasing in the stock market capitalization. This allows us to characterize two extremes in how institutional trade sizes scale across stocks. To facilitate discussion, assume the following scaling relationships across stocks (and omitting subscript t):

$$\left|\mu_{i,\text{unspanned}}\right| \sim M_i^{-\beta_\mu}$$
$$\left|\sigma_{i,\text{unspanned}}\right| \sim M_i^{-\beta_\sigma}$$

In Appendices A.2 and A.3, we empirically estimate that $\beta_\mu \approx 0.2$ and $\beta_\sigma \approx 0.15$. Intuitively, larger stocks have lower variation in expected returns and also lower volatilities. Then, we can characterize two extreme cases:

- **Infinitesimal institutions.** For institutions that are very small ($A \downarrow 0$), $\lambda_i \approx 0$, so we are back to the tradition MV portfolio choice:

$$w_i - w_{i,b} = \frac{\mu_{i,\text{unspanned}}}{\gamma \sigma^2_{i,\text{unspanned}}}$$
$$\Rightarrow |w_i - w_{i,b}| \sim M_i^{-\beta_\mu + 2\beta_\sigma}$$

- **Large institutions.** For very large institutions ($A \uparrow \infty$), TC dominates so we can ignore

volatilities:

$$w_i - w_{i,b} = \frac{\mu_{i,\text{unspanned}}}{\lambda_i}$$
$$\Rightarrow |w_i - w_{i,b}| \sim M_i^{1-\beta_\mu}$$

Of course, reality should lie between the two extremes. We now test these predictions using institutional holding and trading data.

4 Empirical Tests

4.1 Data

To study how institutions trade, we use the quarterly holdings of mutual funds (S12) and 13F institutions (S34) from Thomson Reuters. For the former, we merge it with the CRSP mutual fund database using the MFLINKS of Wermers (2000). We obtain prices, shares outstanding, and trading volume of stocks from CRSP, and we merge it with stock characteristics from Y. Chen and Zimmermann (2022).

Because we are interested in the active positions of institutions, we take care to remove passive funds from our sample. For the S12 mutual fund sample, we remove ETFs, those designated as index funds by CRSP, or those with the word "index" in its name (Dannhauser and Pontiff, 2024). We also require the trailing annual active turnover—defined as |buy dollar amount − sell dollar amount|—to be no lower than 50% of its portfolio and the active share to be no lower than 40%. We impose the same filter on turnover and active share on the S34 institutions. Finally, we avoid overly small institutions by requiring a minimum of 20 stock positions and 10 million USD assets under management.

Table 1 provides summary statistics. Panel A describes the institutions in our sample. For mutual funds, we start from 124 funds at the beginning of the sample and ended with 375. On average, each fund trades around 100 stocks per quarter, and the number of trades approximately doubled over our sample periods. Because we have restricted attention to active funds, the average annual turnover is around 150% - 200% and average active share is around 50%. There is large dispersion in fund sizes. For instance, in the last period, the 95% fund has 7,778 million assets under management (AUM) which is 370 times larger than the 5% fund which only manages 21 million USD. The sample for 13F institutions is larger, starting from 572 institutions at the start of the sample and ending with 1,218 institutions. 13F institutions tend to be larger than mutual funds and hold more stocks, and there is also large variation in AUM across institutions. Panel B describes characteristics of stocks held by mutual funds and institutions. There is also large variation in stock sizes. For instance, in the last period, the 95% stock has a market capitalization of 41,275 million USD which is 616 times that of the 5% stock.

4.2 Testing model implications

We compute two measures of active positioning in stocks. The first is the active deviation from benchmark weights, akin to the definition in Cremers and Petajisto (2009):

$$\text{ActivePosition}_{i,j,t} \equiv |w_{i,j,t} - w^{\text{mkt}}_{i,j,t}|$$

where $w_{i,j,t}$ is the weight of stock i in the portfolio of institution j at the end of quarter t. The benchmark is defined as $w^{\text{mkt}}_{i,j,t} = \frac{M_{i,t}}{\sum_{k \in \mathcal{N}_{j,t}} M_{k,t}}$, where $M_{i,t}$ is the market capitalization of stock i and $\mathcal{N}_{j,t}$ is the set of stocks held by institution j. In other words, we assume that the benchmark is the value-weighted version of all stocks currently held by j.

Panel A: Institutions								
					AUM ($m)			
	N inst	N trades	Turnover	Active share	Mean	5%	50%	95%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Mutual funds (S12)				
1991-1997	124	73	203.2%	50.8%	611	30	263	2,335
1998-2004	308	79	210.8%	50.6%	1,227	24	274	4,767
2005-2011	368	102	182.9%	48.8%	1,345	21	280	5,261
2012-2017	358	117	148.9%	48.1%	1,846	24	363	6,409
2018-2023	375	143	141.9%	48.7%	2,131	21	363	7,778
				13F institutions (S34)				
1991-1997	572	224	137.8%	57.7%	2,362	83	579	10,844
1998-2004	810	236	165.4%	59.9%	3,654	87	560	14,791
2005-2011	1,142	229	179.1%	60.0%	3,512	69	510	14,096
2012-2017	1,208	226	163.7%	58.5%	3,719	57	505	14,983
2018-2023	1,218	257	156.8%	58.2%	4,517	49	503	17,238
Panel B: Stocks								
					Market cap ($m)			
	N stocks	Volume	Vol	Vol (ex-ind)	Mean	5%	50%	95%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1991-1997	5,712	2,443	57.6%	56.9%	933	8	101	3,697
1998-2004	5,754	10,076	62.9%	60.1%	2,309	11	189	7,989
2005-2011	4,498	23,543	50.3%	44.9%	3,477	20	377	14,422
2012-2017	3,607	32,428	39.1%	35.0%	6,032	26	717	24,739
2018-2023	3,452	61,512	49.4%	42.2%	10,853	67	1,285	41,275

Table 1. Summary statistics
We report time-series averages of cross-sectional statistics by time periods. Panel A examines the institutions and the columns report, respectively, the number of institutions, number of stock position changes per quarter, annualized active turnover (in excess of that implied by flow-induced trading), active share, and moments for assets under management (AUM). We separately report results for mutual funds (S12) and 13F institutions (S34). Panel B examines the stocks held by these institutions and the columns report, respective, the number of stocks, average annual trading volume (in million USD), annualized total return volatility, annualized volatility of industry-adjusted returns, and moments of market capitalization.

One drawback of this approach is that we have to make assumptions about the benchmark. To alleviate this concern, we also simply measure the active trades, adjusted for fund flows, as in Section 2.1

$$\text{Trade}_{i,j,t} = \frac{(s_{i,j,t} - s_{i,j,t-1}) \cdot p_{i,t-1}}{\text{AUM}_{j,t-1}} - \underbrace{f_{j,t} \cdot w_{i,j,t-1}}_{\text{flow-induced trading}}.$$

As long as the benchmark does not change significant from quarter to quarter, it is "differenced out" in this trade-based specification.

Full sample results. To investigate how active position sizes scale with stock market capitalization, we start with a panel regression of:

$$\log\left(|y_{i,j,t}|\right) = a_{j,t} + b \cdot \log(M_{i,t-1}) + c' x_{i,j,t-1} + \epsilon_{i,j,t}$$

where $y_{i,j,t}$ can be either portfolio share change induced by active trading ($\Delta w_{i,j,t}$) or position deviation from benchmark ($w_{i,j,t-1} - w^b_{i,j,t-1}$), and $x_{i,j,t-1}$ is a vector of stock characteristics controls including lagged portfolio share $w_{i,j,t-1}$, lagged benchmark portfolio share $w^b_{i,j,t-1}$, and lagged Fama-French six characteristics. By including institution-quarter fixed effects $a_{j,t}$, we only use the cross-section variation of positions within institution-quarter. Standard errors are clustered by institution and quarter.

The regression results are shown in Table 2. Columns (1) and (3) examine how mutual fund active positions scale with log market capitalization. For each unit increase in log market capitalization, log active position sizes increase by 0.328 to 0.428, depending on the active position measure used. Columns (2) and (4) add log, linear, and quadratic controls of return volatility estimated using the previous 63 days, and the results indicate that controlling for volatilities have little effect on the main relationship. In columns (5) and (7), we examine 13F institutions and find even stronger evidence of positions scaling with stock market capitalization, and that controlling for volatilities have little effect. The relationships are both economically and statistically significant.

Sample:	Mutual funds (S12)				13F institutions (S34)			
Dependent variable:	$\log(\lvert\Delta w_{i,j,t}\rvert)$		$\log(\lvert w_{i,j,t-1} - w^b_{i,j,t-1}\rvert)$		$\log(\lvert\Delta w_{i,j,t}\rvert)$		$\log(\lvert w_{i,j,t-1} - w^b_{i,j,t-1}\rvert)$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\log(\text{MktCap}_{i,t-1})$	0.328^{***}	0.335^{***}	0.428^{***}	0.417^{***}	0.408^{***}	0.426^{***}	0.633^{***}	0.624^{***}
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.010)	(0.010)
$\log(\sigma_{i,t-1})$		0.386^{***}		0.058		0.503^{***}		0.060^{*}
		(0.050)		(0.040)		(0.060)		(0.030)
$\sigma_{i,t-1}$		-0.921^{***}		-0.480^{***}		-0.673^{***}		-0.376^{***}
		(0.180)		(0.160)		(0.220)		(0.130)
$\sigma^2_{i,t-1}$		0.165^{***}		0.044		-0.108		0.006
		(0.060)		(0.070)		(0.070)		(0.040)
Inst × Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y
Other controls	Y	Y	Y	Y	Y	Y	Y	Y
Obs	4,121,228	4,121,228	4,121,228	4,121,228	23,953,466	23,953,466	23,953,466	23,953,466
R^2	0.621	0.622	0.601	0.601	0.361	0.362	0.692	0.692
Marginal R^2	0.022	0.022	0.061	0.062	0.029	0.030	0.155	0.155

Table 2. Scaling relationship: Full sample
We use panel regressions to estimate the relationship between active positions ($y_{i,j,t}$) on stock market capitalization:

$$\log\left(\lvert y_{i,j,t}\rvert\right) = a_{j,t} + b \cdot \log(\text{MktCap}_{i,t-1}) + c' x_{i,j,t-1} + \epsilon_{i,j,t}$$

where $y_{i,j,t}$ can be either portfolio share change induced by active trading ($\Delta w_{i,j,t}$) or position deviation from benchmark ($w_{i,j,t-1} - w^b_{i,j,t-1}$). All regressions control for institution × time fixed effects, as well as lagged position $w_{i,j,t-1}$, lagged benchmark position $w^b_{i,j,t-1}$, and Fama-French six characteristics. The even columns control for return volatility and its nonlinear transforms. Standard errors are clustered by institution and quarter. The first four columns use the mutual fund sample while the last four columns use the 13F institution sample. The last row reports the marginal R^2 of market capitalization.

Heterogeneity across institutions. Section 3.2 predict that the trade magnitude – stock size scaling relationship will differ by fund size and the liquidity of the investment universe. We test this prediction in this section.

In each cross-section, to study the effect of fund sizes, we sort institutions into five bins based on lagged AUM. To study the effect of the universe illiquidity of stock universe, we also sort institutions into five bins based on the negative of the average market capitalization of stocks in its portfolio. We then augment the regression in Table 2 by interacting $\log(\text{MktCap}_{i,t-1})$ with these bin indicators.

The results reported in Table 3 are consistent with theoretical predictions. Consider column (1)

which examines how mutual fund trading scaling relates to fund AUM. The trade size-market cap scaling coefficient is 0.24 for funds in the bottom bin based on AUM, and it rises monotonically to 0.24 + 0.20 = 0.44 for those in the top bin. In column (2), we further add interactions with bins by holdings illiquidity, and also find a monotonic relationship there. For funds in the bottom bin for both AUM and holding illiquidity, the trade size scaling coefficient is 0.06, and this rises to 0.06 + 0.20 + 0.33 = 0.59 for those where both are in the top bins. The other specifications show that the finding is robust to using alternative measures of active positions and to using 13F data.

5 Conclusion

This paper documents a robust stylized fact: discretionary institutional trading volumes scale monotonically with stock market capitalization. We show that classical portfolio choice models, such as mean-variance optimization and other variants, fail to account for this stylized fact.

We propose that the missing link is the internalization of transaction costs (TCs). When the classical framework is augmented with a realistic price impact friction, the model not only aligns with the observed cross-sectional scaling of trading volume but also yields new, verifiable predictions. The empirical success of these predictions—specifically regarding how trading patterns vary with institution size and asset liquidity—reaffirms that large investors are not price-takers, but rather internalize the TC incurred by their own trades.

Sample:	Mutual funds (S12)				13F institutions (S34)			
Dependent variable:	$\log(\lvert\Delta w_{i,j,t}\rvert)$		$\log(\lvert w_{i,j,t-1} - w^b_{i,j,t-1}\rvert)$		$\log(\lvert\Delta w_{i,j,t}\rvert)$		$\log(\lvert w_{i,j,t-1} - w^b_{i,j,t-1}\rvert)$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\log(\text{MktCap}_{i,t-1})$	0.24^{***}	0.06^{***}	0.36^{***}	0.04^{**}	0.21^{***}	0.07^{***}	0.39^{***}	0.23^{***}
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
$\log(\text{MktCap}_{i,t-1}) \times$ AUM bin 2	0.00	0.00	−0.00	−0.01	0.01	−0.01	0.05^{***}	0.02^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
$\log(\text{MktCap}_{i,t-1}) \times$ AUM bin 3	0.08^{***}	0.06^{***}	0.04^{*}	0.02	0.05^{***}	−0.01	0.10^{***}	0.04^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
$\log(\text{MktCap}_{i,t-1}) \times$ AUM bin 4	0.13^{***}	0.12^{***}	0.08^{***}	0.06^{***}	0.11^{***}	0.01	0.18^{***}	0.10^{***}
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
$\log(\text{MktCap}_{i,t-1}) \times$ AUM bin 5	0.20^{***}	0.20^{***}	0.14^{***}	0.14^{***}	0.35^{***}	0.18^{***}	0.35^{***}	0.23^{***}
	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
$\log(\text{MktCap}_{i,t-1}) \times$ Illiq bin 2		0.03^{***}		0.10^{***}		0.04^{***}		0.09^{***}
		(0.01)		(0.01)		(0.01)		(0.01)
$\log(\text{MktCap}_{i,t-1}) \times$ Illiq bin 3		0.10^{***}		0.24^{***}		0.09^{***}		0.15^{***}
		(0.01)		(0.01)		(0.01)		(0.01)
$\log(\text{MktCap}_{i,t-1}) \times$ Illiq bin 4		0.22^{***}		0.38^{***}		0.17^{***}		0.22^{***}
		(0.02)		(0.02)		(0.01)		(0.01)
$\log(\text{MktCap}_{i,t-1}) \times$ Illiq bin 5		0.33^{***}		0.50^{***}		0.41^{***}		0.33^{***}
		(0.03)		(0.03)		(0.02)		(0.01)
Inst × Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y
Other controls	Y	Y	Y	Y	Y	Y	Y	Y
Obs	4,121,228	4,121,228	4,121,228	4,121,228	23,953,466	23,953,466	23,953,466	23,953,466
R^2	0.624	0.628	0.603	0.615	0.368	0.372	0.701	0.705
Within R^2	0.079	0.089	0.304	0.325	0.088	0.094	0.494	0.501

Table 3. Scaling relationship: by fund size and holding illiquidity
We estimate panel regressions on active positions ($y_{i,j,t}$) on log lagged market capitalization of stocks:

$$\log(|y_{i,j,t}|) = a_{j,t} + \sum_{\text{bin } x} b_x \cdot I_{j\in x} \log(\text{MktCap}_{i,t-1}) + c' x_{i,j,t-1} + \epsilon_{i,j,t}$$

where $y_{i,j,t}$ can be either portfolio share change induced by active trading ($\Delta w_{i,j,t}$) or position deviation from benchmark ($w_{i,j,t-1} - w^b_{i,j,t-1}$). We sort institutions into five bins based on either assets under management (AUM) or the illiquidity of holdings, defined as the negative of average market capitalization of stocks held, and interact the main regressor with institution bin indicators. All regressions control for institution × time fixed effects, as well as lagged position $w_{i,j,t-1}$, lagged benchmark position $w^b_{i,j,t-1}$, and Fama-French six characteristics. Standard errors are clustered by institution and quarter. The first four columns use the mutual fund sample while the last four columns use the 13F institution sample. The last row reports the marginal R^2 of the liquidity independent variable.

References

Almgren, Robert, Chee Thum, Emmanuel Hauptmann, and Hong Li, 2005, Direct estimation of equity market impact, *Risk* 18, 58–62.

Barber, Brad M, and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773–806.

Basak, Suleyman, and Anna Pavlova, 2013, Asset prices and institutional investors, *American Economic Review* 103, 1728–1758.

Campbell, John Y, Yeung Lewis Chan, and Luis M Viceira, 2003, A multivariate model of strategic asset allocation, *Journal of financial economics* 67, 41–80.

Campbell, John Y, and Luis M Viceira, 1999, Consumption and portfolio decisions when expected returns are time varying, *Quarterly Journal of Economics* 114, 433–495.

Chaudhry, Aditya, and Carter Davis, 2026, The origins of the factor zoo: Investors weakly substitute across stocks, Working paper.

Constantinides, George M., 1986, Capital market equilibrium with transaction costs, *Journal of Political Economy* 94, 842–862.

Cremers, KJ Martijn, and Antti Petajisto, 2009, How active is your fund manager? a new measure that predicts performance, *Review of Financial Studies* 22, 3329–3365.

Dannhauser, Caitlin D, and Jeffrey Pontiff, 2024, Flow, Working paper.

Davis, Carter, 2025, Classical versus demand system asset pricing, Working paper.

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

Edelen, Roger, Richard Evans, and Gregory Kadlec, 2013, Shedding light on "invisible" costs: Trading costs and mutual fund performance, *Financial Analysts Journal* 69, 33–44.

Frazzini, Andrea, Ronen Israel, and Tobias J Moskowitz, 2018, Trading costs, Working paper.

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

Gârleanu, Nicolae, and Lasse Heje Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 68, 2309–2340.

Gârleanu, Nicolae, and Lasse Heje Pedersen, 2016, Dynamic portfolio choice with frictions, *Journal of Economic Theory* 165, 487–516.

Grinold, Richard C., and Ronald N. Kahn, 2000, *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Money Managers*, second edition (McGraw-Hill, New York).

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.

Keim, Donald B, and Ananth Madhavan, 1995, Anatomy of the trading process empirical evidence on the behavior of institutional traders, *Journal of Financial Economics* 37, 371–398.

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

Lo, Andrew W, and Jiang Wang, 2006, Trading volume: Implications of an intertemporal capital asset pricing model, *Journal of Finance* 61, 2805–2840.

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

Markowitz, Harry, 1952, Portfolio selection, *Journal of Finance* 7, 77–91.

Merton, Robert C., 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373–413.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 867–887.

Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29, 104–147.

Novy-Marx, Robert, and Mihail Velikov, 2019, Comparing cost-mitigation techniques, *Financial Analysts Journal* 75, 85–102.

Pollet, Joshua M, and Mungo Wilson, 2008, How does size affect mutual fund behavior?, *Journal of finance* 63, 2941–2969.

Sammon, Marco, and John J Shim, 2023, Do active funds do better in what they trade?, Working paper.

Sharpe, William F, 1963, A simplified model for portfolio analysis, *Management science* 9, 277–293.

Stevens, Guy V. G., 1998, On the inverse of the covariance matrix in portfolio analysis, *Journal of Finance* 53, 1821–1827.

Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses, *Journal of Finance* 55, 1655–1695.

Y. Chen, Andrew, and Tom Zimmermann, 2022, Open source cross-sectional asset pricing, *Critical Finance Review* 11, 207–264.

Appendix

A Additional Empirical Analysis

A.1 Transaction cost modeling

This section provides further details about transaction cost modeling.

TC functional form. In the paper, we approximate the transaction cost (TC) of trading stock i as $\frac{\lambda_i(\Delta w_i)^2}{2}$, where

$$\lambda_i = \overline{\lambda}\frac{A_j^{\alpha_A}}{M_i^{\alpha_M}} \tag{A.1}$$

For institutional-size trading, there are two components of transaction costs: 1) fixed costs, which consists of half of the bid-ask spread and commissions, and 2) variable price impact costs. The total cost can be written as:

$$\underbrace{\Delta w_i \cdot \phi_i}_{\text{fixed cost}} + \underbrace{\frac{\lambda_i^{\text{PI}}}{2}(\Delta w_i)^2}_{\text{price impact}}$$

where ϕ_i is half of the bid-ask spread plus commissions, and and λ_i^{PI} governs the price impact. The relative importance of the two terms above varies. For large funds in small stocks, the price impact term dominates. For small funds in large stocks, the bid-ask spread term dominates.

For the purpose of understanding scaling relationships, we would like to have a single parameter λ_i that approximates the overall cost:

$$\begin{aligned} \overbrace{\frac{\lambda_i}{2}(\Delta w_i)^2}^{\text{approximation}} &= \overbrace{\Delta w_i \cdot \phi_i + \frac{\lambda_i^{\text{PI}}}{2}(\Delta w_i)^2}^{\text{more accurate}} \\ \Rightarrow \lambda_i &= \underbrace{\frac{2\phi_i}{|\Delta w_i|}}_{\text{fixed cost}} + \underbrace{\lambda_i^{\text{PI}}}_{\text{price impact}} \end{aligned} \tag{A.2}$$

To see how the approximation functional form (A.1) works, imagine two extremes. When trade is sufficiently large relative to stock liquidity, it would be dominated by the price impact component. For this component,prior work has shown that stock-level price impact is well described by approximately a constant price multiplier across all stocks. This means that (corresponding to exponents $\alpha_A = \alpha_M = 1$):

$$\lambda_i^{\text{PI}} = \frac{\lambda_0 A_j}{M_i} \tag{A.3}$$

where λ_0 is two times the multiplier, A_j is the AUM of fund j, and M_i is the market capitalization of stock i. Typical estimates imply approximately $\lambda_0 \approx 4$ (e.g. Gabaix and Koijen, 2022).

For smaller trades, we need to take into account the fixed cost component. We can estimate bid-ask spreads from data. As for commissions, Edelen et al. (2013) estimate that it is roughly as large as half of bid-ask spreads. Given this, we assume that ϕ_i is equal to the full bid-ask spread—half of it comes from incurring bid-ask spread and half of it comes from incurring commissions. To estimate bid-ask spread ϕ_i, we directly use the stock-day-level summaries from WRDS Intraday Indicators database. The dependence of the fixed cost component in "λ space", $\frac{2\phi_i}{|\Delta w_i|}$, also depends on $|\Delta W_i|$—which in turn depends on the trading behavior of funds. The larger are the trades, the smaller the importance of the fixed cost component. In the end, how the fixed cost component depends on M_i and A_j is an empirical question.

The fixed and variable costs incurred by funds. To get a sense of the costs incurred by funds, we start with an empirical exploration. We take all quarterly trades by active mutual funds and merge it with bid-ask spreads. We also follow the discussion above to compute λ_i^{PI} for each stock. In each cross-section, we sort funds into 5 bins by AUM and stocks into 5 bins based on market capitalization.

Table A.1 reports summary statistics by the 5×5 bins. Panels (a) and (b) show that, across the sort five bins, fund AUM varies from around 40 million to 8 billion, and stock market capitalization varies from around 150 million to 50 billion. Panel (c) shows the variation in bid-ask spread: the smallest bin of stocks have bid-ask spreads over one percent while the largest bin has 13 basis points.

Consistent with TC mattering for portfolio choice, Panels (d) through (f) show that large funds trade less and tilt their trading towards larger stocks. Panels (d) and (e) report the magnitude of

	Stock market capitalization bins									
AUM bin	1	2	3	4	5	1	2	3	4	5
	(a) Fund AUM (m USD)					(b) Stock market capitalization (m USD)				
1	44	43	42	42	42	154	439	1,057	2,773	49,914
2	129	132	132	132	132	155	441	1,059	2,820	45,754
3	353	353	353	352	351	160	446	1,059	2,766	45,035
4	905	895	890	895	902	157	441	1,056	2,788	43,429
5	4,014	4,528	5,076	5,661	7,893	162	446	1,064	2,869	42,924
	(c) Bid-ask spread (%)					(d) $\lvert\Delta w_{i,j,t}\rvert$ (%)				
1	1.36	0.63	0.39	0.25	0.13	0.37	0.47	0.54	0.63	0.80
2	1.30	0.62	0.39	0.25	0.13	0.27	0.36	0.44	0.52	0.68
3	1.35	0.62	0.39	0.25	0.13	0.16	0.25	0.33	0.44	0.65
4	1.45	0.64	0.40	0.25	0.13	0.08	0.16	0.23	0.33	0.54
5	1.38	0.65	0.40	0.25	0.13	0.05	0.09	0.15	0.23	0.43
	(e) $\sum_i \lvert\Delta w_{i,j,t}\rvert$ (%)					(f) Fraction of trading volume (%)				
1	4.66	6.89	8.83	11.48	31.10	6.65	9.26	13.03	18.61	52.44
2	2.93	5.04	7.08	10.57	29.76	5.21	8.81	12.42	18.95	54.61
3	1.72	4.46	7.02	10.57	27.51	3.23	7.61	12.38	20.09	56.70
4	1.44	3.48	5.52	8.86	27.21	2.20	5.52	9.90	18.41	63.97
5	0.59	1.59	2.84	6.00	26.02	1.16	3.21	6.80	15.60	73.23

Table A.1. By fund and stock sizes: trading behavior
This table examines how mutual fund trading behavior vary by fund AUM and stock market capitalization. In each cross-section, we sort funds into five bins based on AUM and stocks into five bins based on market capitalization. Panel (a) reports fund AUM. Panel (b) reports stock market capitalization. Panel (c) reports bid-ask spread. Panel (d) reports the magnitude of quarterly trades in portfolio weight terms. Panel (e) reports the total trading volume in portfolio weight terms, and Panel (f) reports the fraction of trading volume in each market capitalization bin.

per-stock (total) portfolio weight changes, and the results indicate that larger funds trade less overall and also increasingly shy away from trading small stocks. This is further shown in Panel (f) which reports the *fraction* of trading volume in different stock capitalization bins by fund AUM. The largest funds concentrate their trading more in large stocks while the smallest funds spread their volume more across stock types.

We estimate the amount of TCs incurred by funds in Table A.2 where the left panels report the per-trade (unit) TCs while the right panels report the total impact on portfolio return. The easiest way to understand this table is to first focus on Panels (c) and (d). Panel (c) reports the unit TC due to the fixed cost component, which is just the bid-ask spread (Panel (c) in Table A.1). Panel (d) reports the total impact on fund return which is equal to the bid-ask spreads multiplied by the

amount of trading in portfolio share terms (Panel (e) in Table A.1). As a concrete example, take the top left corner: the smallest funds in smallest stocks. Each trade incurs 1.36% bid ask spread, and these funds trade a total of 4.66% of their portfolio per quarter. This amounts to a total impact of $1.36\% \times 4.66\% \approx 6.33$ basis points of return impact, which is reported in the top left corner of Panel (d) in Table A.2.

To complete the picture, the top and bottom panels in Table A.2 report total and price impact TCs, respectively. The incurrence of costs differ by fund size. Small funds primarily incur fixed costs, and they incur more costs in small stocks which they trade more frequently than the other funds. In contrast, large funds primarily incur price impact costs, and most of their costs are in large stocks. We estimate that the overall quarterly total TC to be around 20 to 25 basis points, which is broadly consistent with the estimates in Edelen et al. (2013).

Estimate scaling exponents α_M and α_A. We now use the 5×5 binned data above to estimate how λ scales with A and M. For each bin, we compute λ for that bin using the formulas in (A.2) and (A.3) where all the formula inputs are averaged over those bins.[5]. Then, we take logs of (A.1) and add an error term to make it amenable to statistical estimation:

$$\log(\lambda_i) = \log(\overline{\lambda}) + \alpha_A \cdot log(A_i) - \alpha_M \cdot \log(M_i) + \epsilon_i$$

and we cluster standard errors by fund size and stock bins. The results are shown in column (1) of Table A.3. As expected, λ is strongly related to both AUM (A) and stock market cap (M). The R^2 of this regression is 0.94, implying strong explanatory power. As robustness, we repeated the analysis by using $n \times n$ bins where $n = 10,20,...,50$. The results are reported in the remaining columns in Table A.3. While the coefficients vary somewhat, the results do not differ significantly. For the main paper, we take the point estimates in the last column with $\overline{\lambda} = e^{3.66} \approx 38.9$, $\alpha_A = 0.58$, and $\alpha_M = 0.79$.

[5]For instance, we take the average fund AUM as A and average stock market cap as M and plug them into (A.3)

	Stock market capitalization bins									
AUM bin	1	2	3	4	5	1	2	3	4	5
	(a) Unit TC: total (%)					(b) Portfolio TC: total (bp)				
1	1.57	0.72	0.43	0.27	0.13	1	7.33	4.96	3.81	3.13	4.11
2	1.74	0.84	0.50	0.30	0.14	2	5.11	4.24	3.53	3.18	4.03
3	2.04	1.01	0.61	0.36	0.14	3	3.50	4.52	4.28	3.85	3.96
4	2.41	1.28	0.79	0.47	0.16	4	3.47	4.44	4.35	4.15	4.25
5	3.68	2.42	1.81	1.17	0.29	5	2.16	3.85	5.16	7.04	7.60
	(c) Unit TC: fixed (%)					(d) Portfolio TC: fixed (bp)				
1	1.36	0.63	0.39	0.25	0.13	1	6.33	4.32	3.43	2.91	4.06
2	1.30	0.62	0.39	0.25	0.13	2	3.81	3.15	2.76	2.67	3.91
3	1.35	0.62	0.39	0.25	0.13	3	2.32	2.76	2.74	2.68	3.68
4	1.45	0.64	0.40	0.25	0.13	4	2.10	2.23	2.19	2.25	3.63
5	1.38	0.65	0.40	0.25	0.13	5	0.81	1.03	1.13	1.53	3.47
	(e) Unit TC: price impact (%)					(f) Portfolio TC: price impact (bp)				
1	0.21	0.09	0.04	0.02	0.00	1	0.99	0.63	0.38	0.22	0.04
2	0.44	0.22	0.11	0.05	0.00	2	1.29	1.10	0.78	0.52	0.12
3	0.69	0.40	0.22	0.11	0.01	3	1.18	1.76	1.55	1.17	0.28
4	0.95	0.63	0.39	0.21	0.02	4	1.38	2.21	2.16	1.90	0.61
5	2.30	1.78	1.42	0.92	0.16	5	1.35	2.82	4.03	5.52	4.13

Table A.2. By fund and stock sizes: quarterly transaction costs (TCs)
This table estimates the TCs incurred by mutual funds by fund AUM and stock market capitalization. In each cross-section, we sort funds into five bins based on AUM and stocks into five bins based on market capitalization. The left panels report the unit (per-trade) TCs while the right panels report the total drag on fund return—unit TCs $\times$ trading volume in portfolio share terms. Panels (a)-(b) report total TCs, which is broken down into the fixed cost component in Panels (c)-(d) and the price impact cost component in Panels (e)-(f).

A.2 How expected returns scale with stock size

We empirically investigate how (changes in) stock expected returns scale with stock sizes. We first use a characteristics-based return prediction model to investigate the objective relationship. We then use analyst forecast-implied expected returns to investigate the subjective relationship.

Characteristics-based expected return. We follow the literature of cross-sectional stock returns to use stock characteristics to predict monthly returns. We use 15 stock characteristics: accruals, asset growth, beta, book-to-market, gross profitbility, industry momentum, intermediate momentum, 1-year net issuance, 5-year net issuance, momentum, seasonality momentum, net operating

Dependent variable: λ						
Num bins:	5	10	20	30	40	50
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	3.59^{***}	3.61^{***}	3.63^{***}	3.61^{***}	3.64^{***}	3.66^{***}
	(0.19)	(0.14)	(0.12)	(0.09)	(0.08)	(0.07)
log(AUM)	0.52^{***}	0.54^{***}	0.56^{***}	0.57^{***}	0.58^{***}	0.58^{***}
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log(Market Cap)	-0.71^{***}	-0.74^{***}	-0.76^{***}	-0.78^{***}	-0.79^{***}	-0.79^{***}
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Obs	25	100	400	900	1,600	2,500
R^2	0.94	0.94	0.95	0.95	0.95	0.95

Table A.3. TC parameter λ scaling
We use panel regressions to estimate the relationship between the all-in price impact coefficient λ and fund AUM and stock market capitalization. Please see Appendix A.1 for implementation details.

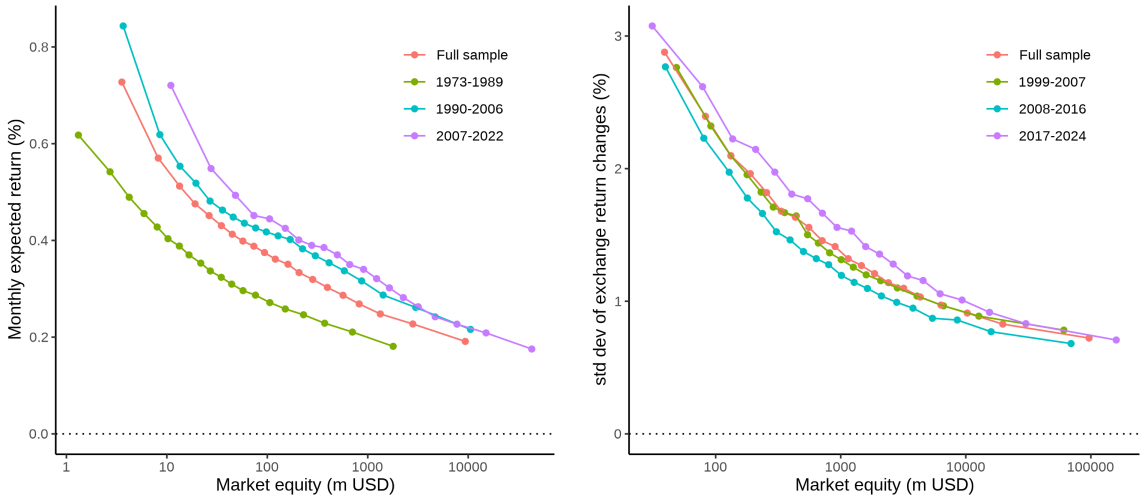
assets, realized volatility, short-term reversals, and size. The characteristics are downloaded from Y. Chen and Zimmermann (2022) and transformed to uniform distributions between -0.5 and 0.5 in the cross-section. We merge these stock characteristics with monthly stock returns from CRSP, and the sample spans 1963 to 2022.

To make the resulting return prediction realistic, we use a walk-forward procedure to estimate the model and predict out of sample. For each year $y = 1973,...,2022$, we use the previous 10 years of data $(y-10, y-9, ..., y-1)$ to estimate a panel regression:

$$r_{i,t} = a + \sum_{k=1}^{K} b_k \cdot x_{i,k,t-1} + \epsilon_{i,t}$$

where $r_{i,t}$ is the excess return of stock i in month t and $\{x_{i,k,t-1}\}_{k=1}^{K}$ are the stock characteristics. We then use this to form predictions for year y.

To gauge how the achievable return predictability vary across stock sizes, in each cross-section, we sort stocks into 20 bins by lagged market capitalization. We form long-short quintile portfolios using the out-of-sample return prediction. The average portfolio returns are plotted against average stock capitalization in Panel (a) of Figure A.1. For the full sample which is plotted by the red line, the long-short quintile portfolio delivers over 0.7% monthly return in the smallest bin, which



(a) Characteristics-based return prediction

(b) Analyst return prediction

Figure A.1. How expected returns scale by stock size.
Panel (a) plots the spread of monthly expected return—based on characteristics-based return predictions—against stock sizes. We use a walk-forward procedure to estimate a return prediction model using a combination of 15 commonly used stock characteristics. We sort stocks into 20 bins based on lagged market capitalization and record the out-of-sample average long-short quintile portfolio returns, which is plotted against the y-axis. Panel (b) examines the variation in analyst price-target implied expected returns. Specifically, we infer the implied monthly return forecast by comparing IBES analysts' consensus 12-month-ahead price targets against prevailing stock prices, and we compute the standard deviation of monthly changes. For both panels, we plot results for the respective full samples and also three equal-length subsamples.

declines monotonically to around 20 basis points for the largest bin. The other lines plot results for four equal-lengthed subsamples and reveal similar findings.

To quantify the scaling relationship, we regress log expected return on log market capitalization of stocks and report results in the first four columns of Table A.4. In the full and subsamples, the scaling coefficient is approximately 0.16. To put magnitudes into perspective, if stock B is 10, 100, or 1,000 time larger than stock A, the result implies that the "alpha opporunities" on it will be $e^{-}0.16 \approx 85\%$, $e^{-0.16\times2} \approx 0.73\%$, or $e^{-0.16\times3} \approx 0.62\%$ as much. We note that the regression R^2 is in a range of 0.96 to 0.99, implying that the log-log specification is reasonably well specified.

Analyst forecast-implied expected return. The characteristics-based return prediction exercise has two possible drawbacks. First, it gauges *objective* expected returns, while the driver of institutional portfolio choice is *subjective* expected returns. Second, it gauges the *level* of return predictability rather than *changes* in return forecasts.

Dependent variable: $\log(\mu)$ or $\log(\Delta\mu)$								
data:	Characteristics-based prediction ($\log(\mu)$)				Analyst price target revisions ($\log(\Delta\mu)$)			
	Full sample	1973-1989	1990-2006	2007-2022	Full sample	1999-2007	2008-2016	2017-2024
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	-0.24^{***}	-0.50^{***}	-0.18^{***}	-0.04	-2.98^{***}	-3.00^{***}	-3.04^{***}	-2.89^{***}
	(0.02)	(0.02)	(0.04)	(0.03)	(0.05)	(0.06)	(0.07)	(0.04)
log(MktCap)	-0.16^{***}	-0.17^{***}	-0.15^{***}	-0.16^{***}	-0.18^{***}	-0.18^{***}	-0.19^{***}	-0.18^{***}
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Obs	20	20	20	20	20	20	20	20
R^2	0.99	0.99	0.96	0.99	0.97	0.96	0.96	0.99

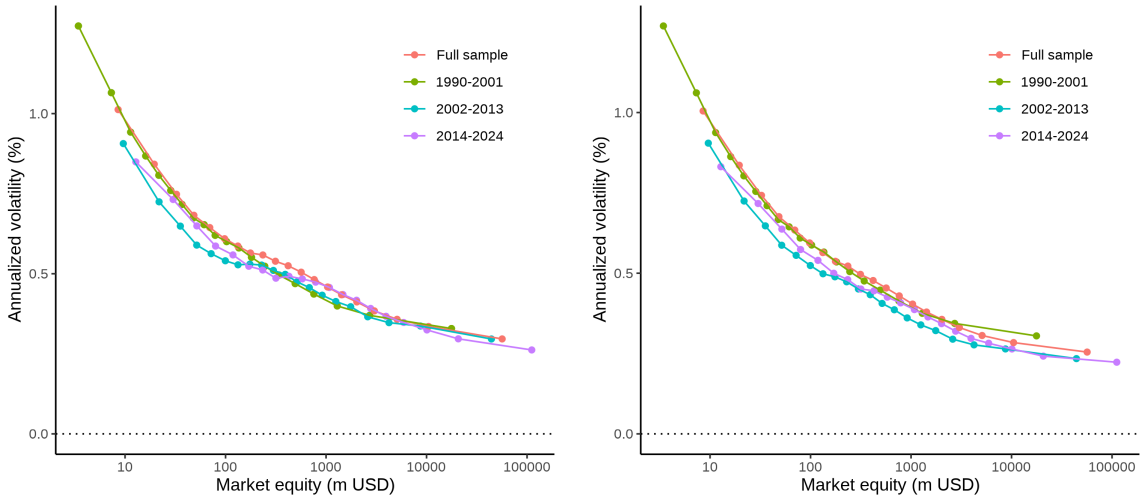
Table A.4. How expected return scales by stock size
We estimate best-fit lines using log-log specifications to the data shown in Figure A.1. The first four columns present results based on characteristics-based return predictions which are plotted in Panel (a) of Figure A.1. The last four columns present results based on analyst price target revisions which are plotted in Panel (b) of Figure A.1.

To this end, we also examine changes in subjective return forecasts implied in equity analyst price targets. We obtain consensus mean analyst 12-month-ahead price targets from IBES which is available from 1999. We impute the implied monthly expected return as

$$\mu_{i,t} = \left(\frac{\text{PriceTarget}_{i,t}}{\text{Price}_{i,t}}\right)^{1/12} - 1.$$

We compute monthly changes as $\Delta\mu_{i,t} = \mu_{i,t} - \mu_{i,t-1}$. To make sure we do not use stale forecasts, we require the price target to have changed from month $t-1$ to t. We then sort the sample cross-sectionally into 20 bins based on lagged stock market cap and compute the standard deviation of $\Delta\mu_{i,t}$ by bin. The results are plotted in Panel (b) of Figure A.1. Relative to the characteristics-based forecasts, analyst-implied forecasts are more volatile. The sample coverage is also tilted more towards larger stocks as microcaps often do not have analyst coverage.

That said, our main interest is in how expected return volatility *scales* with stock sizes, and on this front, the result are remarkably consistent with the earlier ones. Larger stocks have monotonically less volatile expected return variation, and the relationship is reasonably stable across subperiods. In columns (5)-(8) of Table A.4, the regressions reveal that the scaling exponent is in the range of 0.18 to 0.19, which is very similar to that obtained using characteristics-based return predictions.



(a) Total volatility

(b) Residual volatility

Figure A.2. How return volatility scale by stock size.
Similar to Figure A.1, we plot annualized return volatility against stock sizes. Panels (a) and (b) use total and residualized return volatilities, respectively. For both panels, we plot results for the respective full samples and also three equal-length subsamples.

A.3 How volatility scales with stock size

We conduct a similar analysis as in Appendix A.2 to see how stock return volatility scales by stock size. We first compute annualized return volatilities per stock-quarter using daily returns. We compute either total return volatility or residual volatility, defined as the volatility of returns after subtracting Fama-French 12 industry returns.

Figure A.2 plots the volatilities against stock sizes and Table A.5 estimates the regressions. We find that $\sigma \sim M^{-\beta_\sigma}$ with $\beta_\sigma \approx 0.15$.

A.4 Hedging considerations

This section discusses the possibility that the findings in Section 2.1 can be explained by intertemporal hedging considerations. As explained in Lo and Wang (2006), intertemporal hedging leads to a three-fund separation results: in addition to the traditional risk-free and risky portfolio, investors would also have a third portfolio for hedging future investment opportunities. Of course, the identity of this hedging portfolio will in general by investor-specific.

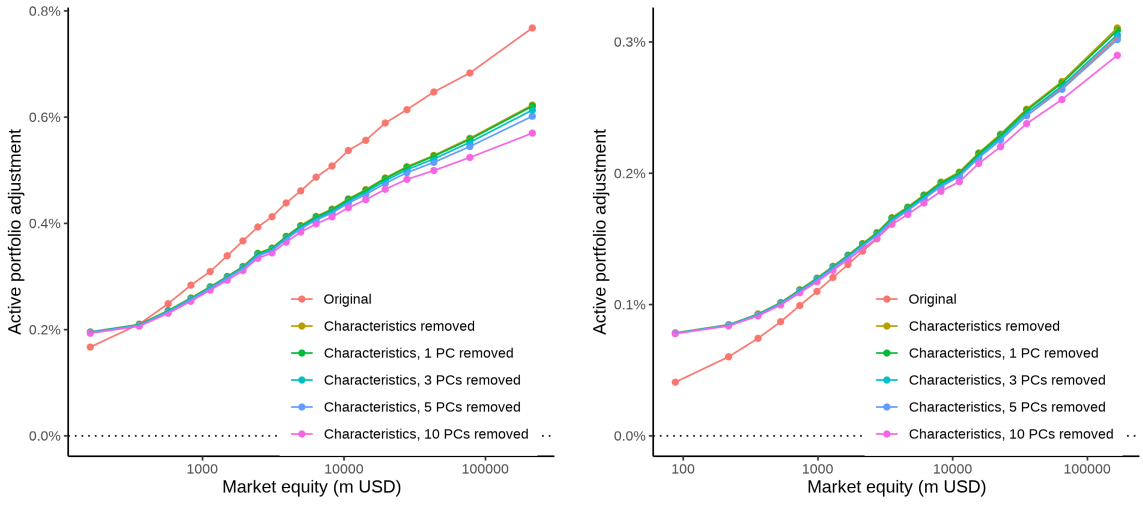
Dependent variable: $\log(\sigma)$								
σ:	Total volatility				Residual volatility			
	Full sample	1990-2001	2002-2013	2014-2024	Full sample	1990-2001	2002-2013	2014-2024
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.19^{***}	0.30^{***}	0.02	0.07^{**}	0.25^{***}	0.32^{***}	0.14^{***}	0.13^{***}
	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
log(MktCap)	-0.14^{***}	-0.16^{***}	-0.13^{***}	-0.13^{***}	-0.16^{***}	-0.18^{***}	-0.16^{***}	-0.16^{***}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Obs	20	20	20	20	20	20	20	20
R^2	0.98	0.96	0.96	0.98	0.98	0.97	0.97	0.98

Table A.5. How return volatilities scales by stock size
We estimate best-fit lines using log-log specifications to the data shown in Figure A.2. The first four columns present results based on total return volatilities plotted in Panel (a) of Figure A.2. The last four columns present results based on residualized return volatilities in Panel (b) of Figure A.2.

Empirical results. We reduce hedging considerations by removing factor components in the trades of each institution j. We take two steps to do so:

- In each period t, use a cross-sectional regression to residualize $\Delta w_{i,j,t}$ against $\{x_{i,k,t-1}\}_{k=1}^{K}$, the set of 15 stock characteristics discussed in Section A.2.
- Further, take the remaining residual and extract institution-specific principal components. We require an institution to have at least 20 quarters of observations to be included. To avoid overfitting, we use even (odd) quarters to estimate PCs and apply the decomposition on odd (even) quarters. Because each institution's set of trades is an unbalanced panel, we use alternating least squares to implement the PCA.

Figure A.3 illustrates the results where we plot the average absolute magnitude of active portfolio shifts ($\Delta w_{i,j,t}$) against 20 bins of stock sorted by size in the cross-section. We observe that the main stylized fact—that large stocks are traded more—is not affected by these operations.



(a) Mutual funds (S12)

(b) 13F institutions (S34)

Figure A.3. Controlling for hedging effects.
These results are similar to that in Figure 1, except that we first remove factors and principal components from trades, as described in Appendix A.4. Panel (a) examines mutual funds while Panel (b) examines 13F institutions.

A summary of Lo and Wang (2006) (LW) LW write down "a very stylized" version of Merton-style ICAPM model in discrete time. In particular, they assume investor i ($i = 1, \ldots, I$) has a myopic but state-dependent utility function, where marginal utility of wealth depends on the dividend for the market portfolio and also an aggregate state variable X_{t+1}, as shown in their Eq. (2). There are also individual state variables Y_t^i and Z_t^i that sum to zero across investors. They assume all shocks $(D_t, X_t, \{Y_t^i, Z_t^i, i = 1, \ldots, I\})$ are IID over time with zero mean, and D_t and X_t are jointly normal, where D_t is the $J \times 1$vector of stock dividends.

They show that this leads investors to hold only two portfolios (two fund separation):[6] the market portfolio and a hedging portfolio to hedge the risk of changes in market conditions, see Eq. (13) in LW:

$$S_t^i = h_{Mt}^i S_M + h_{Ht}^i S_H, \quad \forall i = 1, \ldots, I,$$

where $S_M = (1, \ldots, 1)'$ is the market portfolio, and S_H is the hedging portfolio. These two portfolios are expressed in terms of stock shares and are constant in their model.

Given that investors only hold two portfolios, trading volume also has a two-factor structure. More precisely, in their Proposition 1, LW show that when trading in the hedging portfolio is small

[6] Technically, total portfolios have a three-fund separation: risk-free bond and the two stock funds above.

relative to the market portfolio, "stock turnover" approximately has a two-factor structure: factor 1 is related to market portfolio trading F_{Mt}, and factor 2 is related to hedging portfolio trading, F_{Ht} (see Eq. (16) of LW produced below):

$$\tau_t \approx S_M F_{Mt} + S_H F_{Ht},$$

where τ is stock turnover, and the two factors are defined in Eq. (17) in LW. Crucially, as shown above, a stock's turnover loading on the second factor is proportional to its share weight in the hedging portfolio, S_H.

The empirical implication of the result above is that if we regress stock j's turnover on empirical proxies of F_{Mt} and F_{Ht}, we can recover the percentage share of stock j in the hedging portfolio (as a percentage of its total shares outstanding). LW use two market turnover proxies for F_{Mt} and F_{Ht}: equal-weighted and share-weighted turnover of the market. defined in Eq. (30a)-(30b). They construct estimates of the portfolio weights of the hedging portfolio in Eq. (36).

B Theoretical Derivations

B.1 A simplified case

We show that, in a simplified case of the optimization problem in Section 3.1, considering transaction cost behaves approximately like inflating unspanned variance of stocks.

Consider an investor choosing a portfolio w relative to a benchmark w_b. For analytical simplicity, assume the covariance matrix of returns follows a one-factor structure as in Koijen and Yogo (2019):

$$\Sigma = \sigma_f^2 bb' + D, \quad \text{where } D = \text{diag}(\sigma_{\epsilon,1}^2, \ldots, \sigma_{\epsilon,N}^2)$$

where b is the vector of factor loadings and D contains the idiosyncratic variances. As discussed in Section 3.1, the investor faces quadratic transaction costs governed by a diagonal matrix Λ. The optimization problem is:

$$w = \arg\max_w (w - w_b)'\Delta\mu - \frac{\gamma}{2}(w - w_b)'\Sigma(w - w_b) - \frac{1}{2}(w - w_b)'\Lambda(w - w_b)$$

The first-order condition yields the active weights:

$$w - w_b = (\gamma\Sigma + \Lambda)^{-1} \Delta\mu$$

Substituting the factor structure for Σ, the matrix to be inverted is:

$$\begin{aligned}\gamma\Sigma + \Lambda &= \gamma\sigma_f^2 bb' + \gamma D + \Lambda \\ &= \gamma\sigma_f^2 bb' + \tilde{D}\end{aligned}$$

where $\tilde{D} = \text{diag}(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_N^2)$ is the diagonal matrix of total idiosyncratic friction, with $\tilde{\sigma}_i^2 = \gamma\sigma_{\epsilon,i}^2 + \lambda_i$. Applying the Sherman-Morrison formula, we obtain:

$$(\tilde{D} + \gamma\sigma_f^2 bb')^{-1} = \tilde{D}^{-1} - \frac{\gamma\sigma_f^2 \tilde{D}^{-1} bb' \tilde{D}^{-1}}{1 + \gamma\sigma_f^2 b' \tilde{D}^{-1} b}$$

The active weight for stock i can then be written explicitly as:

$$w_i - w_{b,i} = \frac{1}{\tilde{\sigma}_i^2} \left[\Delta\mu_i - b_i \left(\frac{\sum_j \frac{b_j \Delta\mu_j}{\tilde{\sigma}_j^2}}{\frac{1}{\gamma\sigma_f^2} + \sum_j \frac{b_j^2}{\tilde{\sigma}_j^2}} \right) \right]$$

Defining the term in the parentheses as the aggregate factor signal $\Delta\mu_f$, the expression simplifies to:

$$w_i - w_{b,i} = \frac{\Delta\mu_i - b_i \Delta\mu_f}{\gamma\sigma_{\epsilon,i}^2 + \lambda_i}$$

The numerator $(\Delta\mu_i - b_i\Delta\mu_f)$ is the unspanned (idiosyncratic) expected return signal. This derivation shows that considering transaction cost amounts to inflating the unspanned variance.

B.2 Portfolio choice with transaction costs: generic case

Consider a portfolio choice problem where an investor chooses among N risky assets, has benchmark weight w_b, and her previous-period positions are the same as the benchmark. Her

portfolio choice problem is:

$$\begin{aligned} w &= \arg\max_{w}(w - w_b)'\Delta\mu - \frac{\gamma}{2}(w - w_b)'\Sigma(w - w_b) - \frac{1}{2}(w - w_b)'\Lambda(w - w_b) \\ &= w_b + (\gamma\Sigma + \Lambda)^{-1}\,\Delta\mu \end{aligned}$$

As discussed in Section 3.1, we assume that Λ is a diagonal matrix. Similar to Appendix A, we can subdivide the to-be-inverted matrix as:

$$\gamma\Sigma + \Lambda = \begin{bmatrix} \gamma\Sigma_{-N,-N} + \Lambda_{-N,-N} & \gamma\Sigma_{-N} \\ \gamma\Sigma'_{-N} & \gamma\sigma_N^2 + \lambda_N \end{bmatrix}$$

Using the block diagonal matrix formula, the inversion is given by:

$$(\gamma\Sigma + \Lambda)^{-1} = \begin{bmatrix} (\gamma\Sigma_{-N,-N} + \Lambda_{-N,-N})^{-1} + \tau_N\beta_{-N}\beta'_{-N} & -\tau_N\beta_{-N} \\ -\tau_N\beta'_{-N} & \tau_N \end{bmatrix}$$

where we define the transaction-cost adjusted hedge ratios β_{-N} and the adjusted unspanned precision τ_N as:

$$\begin{aligned} \beta_{-N} &= (\gamma\Sigma_{-N,-N} + \Lambda_{-N,-N})^{-1}(\gamma\Sigma_{-N}) \\ \tau_N &= \left((\gamma\sigma_N^2 + \lambda_N) - \gamma\Sigma'_{-N}\beta_{-N}\right)^{-1} \end{aligned}$$

To gain further intuition, let $\beta_{-N}^{MV} = \Sigma_{-N,-N}^{-1}\Sigma_{-N}$ be the standard mean-variance hedge ratios. By applying the Woodbury Matrix Identity to β_{-N}, we can see how the diagonal costs $\Lambda_{-N,-N}$ suppress the hedging activity:

$$\beta_{-N} = \left(I + \frac{1}{\gamma}\Lambda_{-N,-N}\Sigma_{-N,-N}^{-1}\right)^{-1}\beta_{-N}^{MV}$$

By extracting the last row of the matrix inversion, the optimal active weight for asset N is given by:

$$\begin{aligned} w_N - w_{b,N} &= -\tau_N\beta'_{-N}\Delta\mu_{-N} + \tau_N\Delta\mu_N \\ &= \tau_N(\Delta\mu_N - \beta'_{-N}\Delta\mu_{-N}) \end{aligned}$$

To provide a clearer intuition regarding the scaling across stocks, we divide both the numerator and the denominator by γ. Let $\hat{\lambda}_i = \lambda_i/\gamma$ be the "variance-equivalent" transaction cost. The expression becomes:

$$w_N - w_{b,N} = \frac{\Delta\mu_N - \beta'_{-N}\Delta\mu_{-N}}{\gamma\left[(\sigma_N^2 + \hat{\lambda}_N) - \Sigma'_{-N}\beta_{-N}\right]}$$

Using the definition of the unspanned return from Equation (7) in the text[cite: 33], we can write the final explicit expression as:

$$w_N - w_{b,N} = \frac{1}{\gamma} \cdot \frac{\Delta\mu_{N,unspanned}(\beta_{-N})}{\sigma_N^2 + \hat{\lambda}_N - \Sigma'_{-N}\beta_{-N}}$$

where the "adjusted" unspanned return and the denominator depend on the cost-mitigating hedge ratios:

$$\beta_{-N} = (\Sigma_{-N,-N} + \hat{\Lambda}_{-N,-N})^{-1}\Sigma_{-N}$$