Bottom-Up Capacity Constraints and the Limits of Anomaly Profitability

Álvaro Cartea [*†] Mihai Cucuringu [*‡§] Qi Jin [*‡] Jiexiu(Victoria) Zhu [*‡¶]

Click here for latest version

Abstract

We propose a new bottom-up framework to evaluate the implementable value of trading strategies. Our approach measures profitability in terms of dollar trade volumes and returns, thereby quantifying feasible trading capacity at the stock level. Rather than focusing solely on statistical measures of predictability, we explicitly incorporate capacity constraints that limit the scale at which signals can be deployed. Applying the framework to a broad set of asset pricing anomalies and return prediction models, we find that the realized Sharpe ratio of anomaly-based strategies declines substantially once stock-level capacity constraints are taken into account. Although many anomalies and machine-learning-based models appear highly predictive in statistical terms, their predictive power is disproportionately concentrated in illiquid stocks with limited tradable volume and therefore restricts scalable implementation. While returns are further reduced when trading costs are incorporated, capacity constraints represent the primary limitation on the feasibility of anomaly-based strategies. Our findings reveal a divergence between predictability and profitability and highlight the importance of evaluating trading strategies based on implementable value and achievable dollar returns under realistic capacity limits.

Keywords: Asset pricing anomalies, capacity constraints, implementability, trading frictions

JEL classification: G12, G11, G14

[*]Oxford-Man Institute of Quantitative Finance, University of Oxford
[†]Mathematical Institute, University of Oxford
[‡]Department of Statistics, University of Oxford
[§]Department of Mathematics, University of California, Los Angeles
[¶]Corresponding author jiexiu.zhu@hertford.ox.ac.uk

We would like to thank Nan Chen, Olga Kolokolova, Roman Kozhan, Yingying Li, José Penalva, and Mao Ye for many helpful comments. We also benefited from comments made by seminar participants at University of Oxford and Lancaster University.

1 Introduction

Over the past decades, asset pricing research has identified an extensive array of stock return predictors and model-based strategies, many reporting substantial improvements in predictive performance. However, predictive accuracy alone does not guarantee investment success. Most studies evaluate economic performance as percentage returns on a \$1 investment.[1] While such measures capture relative gains, they overlook a critical question: how much can one sensibly invest? In practice, scaling a theoretically profitable strategy to institutional size is not trivial. When a strategy targets small and illiquid stocks, its economically implementable value is severely limited.

We propose a bottom-up framework that quantifies how much investors can feasibly trade while accounting for stock-level capacity constraints. We proxy trading capacity with each stock's average daily volume (ADV) and model trade size as a function of the stock's predicted return together with its ADV. Rather than allocating a fixed amount of capital top-down across the cross section, our bottom-up approach directly estimates the tradable dollar volume of each individual stock.

Crucially, our framework measures profitability directly in dollar return, which links predictive signals to their implementable economic value. Although percentage returns are the standard performance metric for comparing strategies across time, this normalized measure does not convey information about the magnitude of profits or the feasible scale of trading. Our framework recovers percentage return as the ratio of dollar return to traded dollar volume. In this way, we jointly capture the relative profitability and the implementable scale of a trading strategy, providing a more comprehensive evaluation of strategy performance.

We apply the framework to re-evaluate the performance of 126 cross-sectional return predictors. Each predictor is fitted with a univariate regression model to forecast the market excess returns of stocks. Prior studies attribute the deterioration in anomaly performance to statistical bias, post-publication decay (McLean and Pontiff, 2016), and trading costs (Novy-Marx and Velikov, 2015; Chen and Velikov, 2023). To disentangle

[1] This convention is used in early cross-sectional predictor studies such as Basu (1977), Banz (1981), and Jegadeesh and Titman (1993). It is also used in more recent predictor discovery papers including Park (2005) and Novy-Marx (2013); in works examining the performance of return prediction models such as Gu et al. (2020) and Chen et al. (2024); and in reassessments of existing anomalies including Novy-Marx and Velikov (2015), McLean and Pontiff (2016), Jesen et al. (2023), and Chen and Velikov (2023).

these effects, we divide the sample into in-sample (IS) and out-of-sample (OOS) periods, where the IS period matches the window used in the original publication. We first assume zero trading costs to isolate the impact of capacity constraints on strategy performance, and then incorporate cost estimates to examine how trading costs further erode profitability.

Capacity constraints alone materially reduce much of the profitability of single-anomaly strategies. More than 60% of single-anomaly strategies generate less than \$1,000 in daily dollar return during the OOS period. Realized Sharpe ratios and return t-statistics also decline substantially relative to the conventional one-dollar portfolio benchmark when tradable volume limits are applied for each stock. The realized returns of fewer than half of the anomalies within their original IS periods are statistically significant, and only 24 of the 126 predictors are significant OOS. Overall, capacity constraints substantially restrict the implementable value and realized profitability of most anomalies.

The results reveal a clear divergence between predictability and profitability. Many well-known predictors, such as short-term reversal, exhibit strong OOS predictive power yet produce low realized Sharpe ratios and negligible dollar returns when capacity constraints are accounted for. By contrast, the few "surviving" anomalies are mostly related to profitability and external financing. Despite their weaker predictability, these signals tend to operate in larger, more liquid firms and therefore admit larger trading capacity and dollar return. This finding challenges the common presumption that strong predictive performance implies greater economic significance, and it highlights the risk of relying on portfolio percentage returns as an indication of profitability when trading on these anomalies.

To further examine the divergence between predictability and profitability, we analyze results across stock-size quintiles. Small-cap stocks generally exhibit much stronger predictability than large-cap stocks with higher percentage returns, yet their limited tradable volume constraints the amount of capital that can be deployed. In contrast, most dollar returns originate from large-cap stocks, where predictability is weaker and less stable over time but with higher capacity. These patterns illustrate how predictive strength and implementable scale jointly shape economic value.

If most single anomalies suffer from capacity constraints, what about combining them? To address this, we train several commonly used prediction models (including penalized

linear, tree-based, and ensemble models) to aggregate the predictors. Consistent with previous findings (Haugen and Baker, 1996; Gu et al., 2020), these prediction models substantially enhance predictive performance. However, the improvement in profitability is modest. The gains remain concentrated among small-cap stocks, and the improved forecast precision does not lead to an increase in realized investment gains in dollar terms.

Assuming zero trading costs, we find that capacity constraints alone substantially limit the implementable scale and realized returns of anomaly-based strategies. Including trading costs, which are widely documented as the major trading friction in the literature (Frazzini et al., 2012; Novy-Marx and Velikov, 2015; Frazzini et al., 2018; DeMiguel et al., 2020; Chen and Velikov, 2023; Muravyev et al., 2025), further compound these limitations. To provide a comprehensive assessment of strategy performance that incorporates all key frictions, we use two standard proxies to estimate trading costs: the half bid-ask spread (Chen and Velikov, 2023), and price impact costs (Frazzini et al., 2018).

Most prior studies subtract estimated costs directly from gross returns to account for trading costs. We adopt a more realistic approach: we compare expected gross returns with cost estimates and only execute trades that are expected to yield non-negative post-cost returns. When half-spread costs are paid, more than 50% of predicted trades are excluded because their expected gross returns are insufficient to offset cost estimates. This provides an alternative explanation for why some studies conclude that trading costs nearly eliminate positive anomaly returns: by subtracting aggregate trading costs from portfolio returns, they include numerous trades with negative expected net returns that would not be executed in practice.

When imposing capacity constraints and assuming zero trading costs, only 24 of 126 predictors deliver significant OOS percentage returns. Accounting for trading costs reduces this number further: 18 are significant when price impact costs are accounted for, or 9 are significant when half-spread costs are accounted for. These results highlight capacity constraints as the primary limitations that restricts feasible trading volume and, in turn, the corresponding profitability. We emphasize that trading costs should be evaluated jointly with capacity limits, rather than in isolation, to assess the feasibility of trading strategies.

Finally, we demonstrate the practical relevance of our framework with a fund-size case

study. The results are mixed: as investment capital increases, the average daily dollar return rises, but the realized Sharpe ratio declines. The latter pattern is consistent with the well-documented diseconomies of scale in asset management (Grinblatt and Titman, 1989; Perold and Salomon Jr., 1991; Chen et al., 2004; Berk and Green, 2004; Yan, 2008). When the fund is small and nimble, it can invest capital in stocks with attractive expected returns. As the fund grows and capacity constraints start to bind, cash is diverted into a broader set of securities where the additional stocks are less profitable but more scalable. This illustrates that strategies deliver varying benefits depending on the scale at which they are implemented.

Our findings have implications for both research and practice. For research, we show that capacity constraints play a central role in limiting the profitability of anomaly-based strategies. The framework provides a more realistic assessment of their economic value. For practitioners, we offer a systematic way to evaluate implementable scale and achievable profits, which helps portfolio managers and quantitative investors to align the design of strategies with different capacity levels. More broadly, our work bridges the gap between statistical return predictability and practical strategy implementation by showing that overlooking capacity constraints can misrepresent achievable Sharpe ratios and realized profitability.

A large body of literature re-examines the wide spectrum of anomalies documented in the "factor zoo". The predictive power of anomalies often weakens OOS (Harvey et al., 2015), and declines further after publication (McLean and Pontiff, 2016). Our results are consistent with these findings: both the R^2 and the Sharpe ratio of the portfolio decline substantially in the OOS period relative to their IS performance. However, our evidence also shows that the performance of the strategy during the original IS period is already subject to severe capacity constraints. This raises concerns about implementability even before the OOS deterioration.

Most studies that examine market frictions focus on trading costs (Lesmond et al., 2004; Frazzini et al., 2012; Novy-Marx and Velikov, 2015; Frazzini et al., 2018; DeMiguel et al., 2020; Chen and Velikov, 2023; Muravyev et al., 2025). Chordia et al. (2014) shows that improvements in market liquidity enhance market efficiency and reduce anomaly returns. Korajczyk and Sadka (2004), Frazzini et al. (2012), and Novy-Marx and Velikov (2015) also acknowledge limited trade capacity as part of the trade-off between position

size and price impact. However, all of these studies adopt a traditional top-down portfolio construction approach. To the best of our knowledge, our study is the first to implement a bottom-up dollar allocation framework.

Our approach isolates capacity limits as independent constraints and demonstrates that they impose substantial limitations on strategy profitability even before accounting for transaction costs. Moreover, while most analyses of trading costs focus on gross percentage returns, our framework complements this literature by evaluating implementable trading volume and realized profits directly in dollar terms.

Our study also revisits the effectiveness of return prediction models. Prior work shows that combining anomalies improves portfolio performance (Haugen and Baker, 1996; DeMiguel et al., 2020). More recently, the proliferation of return predictors has spurred the use of machine learning techniques in empirical asset pricing. Gu et al. (2020) and Chen et al. (2024) show that advanced learning architectures can substantially improve forecast performance. However, these studies evaluate profitability through portfolios and largely overlook capacity limits. Our paper goes a step further, it examines whether improvements in predictive power lead to tangible gains in dollar return. Much of the increase in predictability is concentrated in small-cap stocks with limited tradable volume. This poses a challenge for models that pursue higher predictive accuracy without considering scalability. Future research should aim to improve predictability in areas where investors can deploy larger amounts of capital.

The remainder of the paper proceeds as follows. Section 2 describes the data and the framework. Section 3 reassesses the profitability of anomaly-based strategies with our framework assuming zero trading costs, while Section 4 accounts for trading costs. Section 5 presents a case study of hedge fund performance with various levels of investment capital. Section 6 concludes. Model details and additional results are in the appendix.

2 Data and Methodology

In Section 2 we first describe the data we use in the analysis to construct predictors and replicate asset pricing anomalies. Next, we introduce the key methodology in our framework and emphasize the distinction between the predictability and profitability evaluation.

2.1 Anomalies Data

The data consist of returns and firm characteristics for common stocks publicly traded on the New York Stock Exchange, American Stock Exchange, and Nasdaq. We exclude micro-cap stocks with prices below \$5. Most of the anomalies are from Chen and Zimmermann (2022), which provides open-source code to reproduce 205 cross-sectional return predictors, along with a comprehensive assessment of their quality.[2] The raw data are from the WRDS online database, spanning January 1930 to December 2023. Market variables — such as prices, volumes, and returns — are sourced from the Center for Research in Security Prices (CRSP) daily and monthly databases. For simplicity, we use the word "anomalies" and "predictors" interchangeably. We restrict our analysis to replicable anomalies classified as "best quality" (score of 1) in the "Signal Replicable Quality" rating in Chen and Zimmermann (2022), and further exclude anomalies with more than 70% missing observations. This screening yields 105 predictors. In addition, we construct 21 daily predictors using CRSP daily data where available, resulting in a final set of 126 anomalies. See Appendix D for a complete list of anomalies.

Standardizing predictor data allows direct comparison of coefficients across anomalies and is essential when applying a ridge penalty. We implement an exponentially weighted average standardization scheme to update each anomaly's sample mean and variance as new data become available. This approach is free from look-ahead bias and adapts to the non-stationary nature of financial data.

Suppose $x_{i,t}$ is the value of an anomaly that we use to predict the returns of stock i at time t. We define a decay coefficient λ according to a half-life of h, where $\lambda = \exp\{\log(0.5)/h\}$. At each time t, with n_t new observations, the exponentially weighted mean μ_t and variance σ_t^2 of anomaly x_t are updated as

$$\mu_t = \frac{\lambda\, n_{1:t-1}\, \mu_{t-1} \;+\; n_t\, \bar{x}_t}{\lambda\, n_{1:t-1} \;+\; n_t} \quad \text{and} \quad \sigma_t^2 = \frac{\lambda\, n_{1:t-1}\, \sigma_{t-1}^2 \;+\; n_t\, s_t^2}{\lambda\, n_{1:t-1} \;+\; n_t}\,,$$

where $n_{1:t-1}$ is the total number of observations until time $t-1$, and $\bar{x}$ and s_t^2 are the sample mean and the sample variance for the new observations at time t, respectively.

[2]Our paper uses the Aug 2024 data release, see `https://www.openassetpricing.com/`.

Therefore, the standardized value of the anomaly is

$$z_{i,t} \;=\; \frac{x_{i,t} - \mu_t}{\sigma_t}\,,$$

which we use in our model to predict stock returns.

2.2 Market Excess Return

In this paper, we predict the market excess return of each stock. The market excess return of an asset is the residual from the capital asset pricing model (CAPM),

$$R_{i,t} = \beta_i\, R_t^M \;+\; \epsilon_{i,t}\,, \tag{1}$$

where $R_{i,t}$ is the percentage raw return of asset i at time t, R_t^M is the return of a broad-based market portfolio (in our analysis, the value-weighted S&P 500), and β_i is the market beta of asset i. The residual term $\epsilon_{i,t}$ is the market excess component of the return, which is also often referred to as the market-adjusted or CAPM-adjusted return, see, e.g., Rasekhschaffe and Jones (2019).

Online Bayesian Regression. The financial market is dynamic and often undergoes regime shifts. Therefore, we employ a standard online Bayesian regression (OBR) to estimate the CAPM beta and residual returns. OBR uses regression analysis and Bayes' theorem to update the posterior distribution of the regression parameter β based on both prior information and incoming data. Unlike traditional fixed-window regressions that assume static relationships, which may be unreliable in evolving market conditions, OBR updates parameter estimates sequentially, thereby reducing the computational burden of refitting the model on the entire dataset at each time step. This approach allows the model adapt to regime shifts and changes in asset risk profiles more quickly. In particular, OBR allows our CAPM model to update its coefficients daily, whereas fixed-window regression updates coefficients only when the model is refitted; see Appendix A.1 for more details.

To implement OBR in (1), we assign a prior distribution to β_i for each asset $i = 1, \ldots, N$. At each time step t, we observe the return pair $(R_{i,t}, R_t^M)$ and update the posterior distribution of β_i. Let $\widehat{\beta}_{i,t-1}$ and $\widehat{\Sigma}_{i,t-1}$ denote the posterior mean and variance

of β_i from the previous step, which are the prior parameters for the current update. The corresponding precision matrix is $P_{i,t-1} = \widehat{\Sigma}_{i,t-1}^{-1}$.

When estimating the market beta for each stock, the regression model is univariate with the market return R_t^M as the explanatory variable, and the design matrix for asset i at time t is a scalar. To update the precision matrix, we scale the prior precision matrix by the exponential decay factor λ, which controls the weight of older observations, and add the contribution from the new observation, so

$$P_{i,t} \;=\; \lambda\, P_{i,t-1} \;+\; \left(R_t^M\right)^2 ,$$

where, as above, $\lambda = \exp\{\log(0.5)/h\}$.

The updates of the posterior mean and covariance

$$\widehat{\beta}_{i,t} \;=\; P_{i,t}^{-1}\Big(\lambda\, P_{i,t-1}\,\widehat{\beta}_{i,t-1} \;+\; R_t^M\, R_{i,t}\Big) \quad \text{and} \quad \widehat{\Sigma}_{i,t} = (P_{i,t})^{-1}$$

are sequential, so we obtain time-varying estimates of the market beta, which adjusts flexibly to new information and reflects the evolving nature of asset sensitivities over time.

With the updated posterior mean for CAPM beta, the estimate of the market excess return for asset i at time $t+1$ is

$$r_{i,t+1} \;=\; R_{i,t+1} \;-\; \widehat{\beta}_{i,t}\, R_{t+1}^M .$$

2.3 Performance Evaluation

When predicting stock market excess returns, we consider

$$r_{i,t} = \mathbb{E}[r_{i,t} \mid F_{t-1}] \;+\; \eta_{i,t}\,, \tag{2}$$

where $\mathbb{E}[r_{i,t} \mid F_{t-1}]$ is the expected excess return of stock i at time t given information available at time $t-1$. We model this conditional expectation as

$$\mathbb{E}[r_{i,t} \mid F_{t-1}] = f^*(X_{i,t-1})\,,$$

where $X_{i,t-1}$ is the predictor vector and $f^*(\cdot)$ is a common predictive function. Recall that stocks are indexed by $i = 1, \ldots, N$, and time (days/ months) by $t = 1, \ldots T$. The function $f^*(\cdot)$ is assumed to be invariant across stocks; therefore, we fit the same model for all stocks. For simplicity, in the subsequent analysis, we refer to $f^*(X_{i,t-1})$ as the expected excess return $\widehat{r}_{i,t}$.

Next, we describe the measures we use to evaluate the performance of market excess return forecasts. Importantly, we make the distinction between predictability of stock returns and tradability of market excess return forecasts, and we emphasize that predictability of stock returns does not equate to economic significance, nor does it guarantee profitability of the corresponding trading strategies.

Conventional Measures and Their Limitations. Here, we consider a weighted version of the out-of-sample R^2

$$R^2 = 1 - \frac{\sum_{(i,t)} w_{i,t} \left(r_{i,t} - \widehat{r}_{i,t}\right)^2}{\sum_{(i,t)} w_{i,t} \left(r_{i,t}\right)^2}, \tag{3}$$

where $w_{i,t}$ denotes the weight assigned to each observation, such that stocks of greater "importance" receive larger weights.

A natural notion of importance arises from trading capacity. Liquid and high-capacity stocks can absorb larger trades with limited price distortion, which tend to be more efficiently priced and exhibit less mispricing (see, e.g., Tarun et al. (2008)). Therefore, it is particularly relevant to assess whether predictive signals retain explanatory power for stocks with larger trading capacity. From a practical standpoint, capacity limits also determine scalability: more capital can be invested easily in stocks with sufficient tradable volume that can accommodate large trades with limited price impact. Consequently, predictive accuracy in liquid stocks with higher capacity carries greater economic value than that in low-capacity stocks where execution frictions and trading constraints can substantially reduce execution feasibility and realized profitability.

We use ADV as a proxy for individual stock trading capacity, because it reflects recent trading activity and is widely used in both academic research and practice as a benchmark for practical liquidity and tradable dollar capacity (Jegadeesh and Wu, 2022). Although capacity can be measured using richer microstructure objects, such as order-book depth or direct estimates of price impact, ADV provides a practical and scalable

proxy that determines how much capital can be deployed in practice. The weights in (3) are $w_{i,t} = \log(\text{ADV}_{i,t})$, where $\text{ADV}_{i,t}$ is the average daily volume traded over past 30 days for stock i. This weighting scheme places greater emphasis on more liquid and high-capacity stocks when evaluating predictive performance.

In addition to the weighted R^2, most studies evaluate the economic contribution of the performance of anomalies at the portfolio level. A key step in the portfolio construction is to determine top-down how total capital is allocated across stocks. For simplicity, much of the literature assumes a unit capital of \$1 (see, e.g., Jegadeesh and Titman (1993); McLean and Pontiff (2016); Chen and Zimmermann (2022)), so stock weights and portfolio returns can be directly expressed in percentage terms.

The standard approach in the literature is to construct long-short portfolios based on either raw predictor values (Jegadeesh and Titman, 1993; Naranjo et al., 1998; Kishore et al., 2008; Loughran and Wellman, 2011), or predicted returns (Jegadeesh, 1990; Gu et al., 2020). These portfolios typically go long in the top quintile of stocks and short in the bottom quintile (Schwert, 2003; McLean and Pontiff, 2016). Recent literature considers optimal portfolio weights that maximize expected utility or the Sharpe ratio of the portfolio (Guijarro-Ordonez et al., 2021; Chen et al., 2024).

In both approaches, the portfolio returns can be approximated by the covariance between the normalized weights and the stock returns

$$\mathbf{w}_t^{*\top} \mathbf{r}_t \;=\; \sum_{i=1}^{N} w_{i,t}^{*}\, r_{i,t} \propto Cov[w_{i,t}^{*}\,,\, r_{i,t}]\,,$$

where $w_{i,t}^{*}$ denotes the normalized portfolio weights derived from signal rankings or utility-maximizing allocations. In long-short strategies, the weight $w_{i,t}^{*}$ typically takes the discrete form as $(+1\,,\,-1)$ to reflect long and short positions. In utility-maximizing strategies, the weight $w_{i,t}^{*}$ corresponds to the continuous optimal allocation across stocks.

We first follow the top-down approach and construct a capacity-weighted portfolio with unit-dollar capital where the return of the portfolio is

$$R_t^P \;=\; \sum_{i=1}^{N} \frac{w_{i,t}\, \widehat{r}_{i,t}\, r_{i,t}}{\sum_{i=1}^{N} w_{i,t}\, |\widehat{r}_{i,t}|}\,. \tag{4}$$

The weight of the stock in the portfolio is consistent with the weighting scheme used in

the computation of R^2 where stock weights are proportional to both the predicted excess return $\widehat{r}$ and to log(ADV).

The corresponding return t-statistic and (annualized) Sharpe ratio of the portfolio are

$$t_H = \frac{\overline{R^P}}{s(R^P)}\sqrt{T} \quad \text{and} \quad \text{Sharpe}_H = \frac{\overline{R^P}}{s(R^P)}\sqrt{252}\,, \tag{5}$$

where $\overline{R^P} = \frac{1}{T}\sum_{t=1}^{T} R_t^P$ is the sample mean of portfolio return, $s(R^P)$ is the sample standard deviation, and the factor $\sqrt{252}$ annualizes the Sharpe ratio (assuming 252 trading days per year).

However, the theoretical returns from either equal-weighted or capacity-weighted portfolios are rarely achievable in practice. In particular, such portfolio constructions typically do not account for the capacity constraints of individual stocks in real-world trading. The portfolio return in (4) is the percentage return of the one-dollar portfolio or the first dollar invested, but investors cannot expect to maintain the same percentage return or Sharpe ratio (5) as investment capital increases.

Institutional capital often involves millions or billions of dollars, and the limited trading capacity of small- and mid-cap stocks generally prevents full implementation of the theoretically optimal portfolio weights.

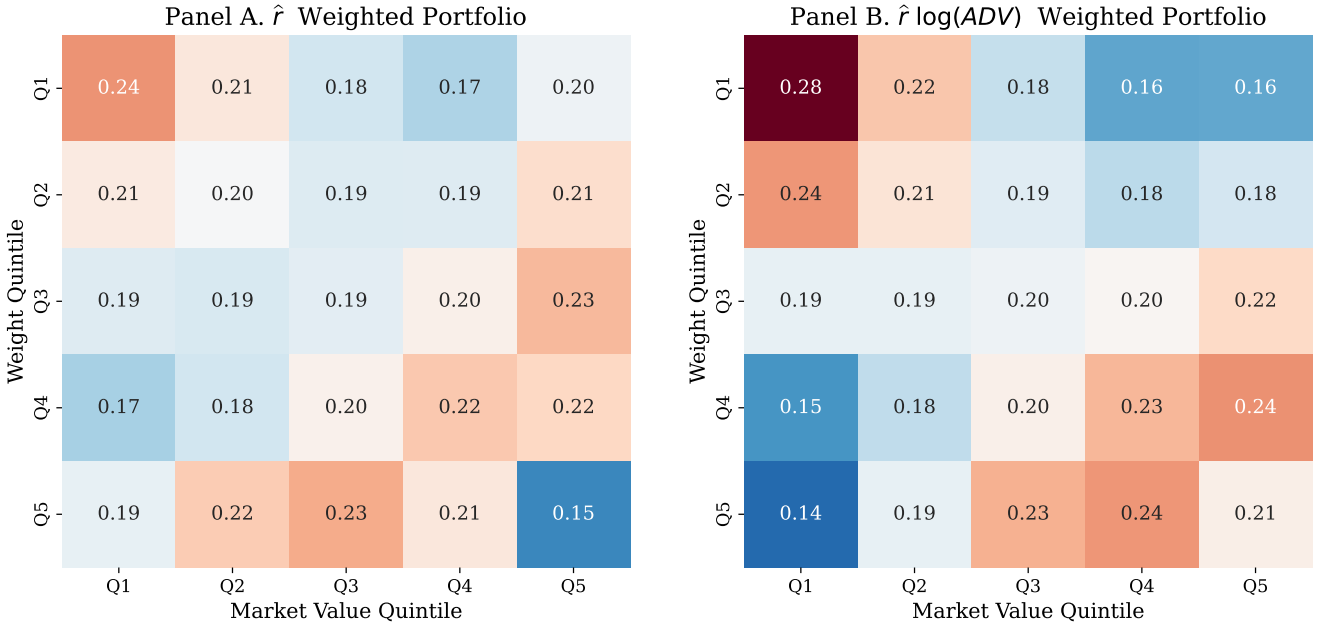
To illustrate the implementation challenge more concretely, we take the one-day momentum anomaly as an example. On each date, we use market value to sort all stocks into five size quintiles. Next, we construct two alternative portfolios: (i) weights are proportional to the predicted excess return $\widehat{r}$ from a univariate regression; and (ii) capacity-adjusted weights $\widehat{r}\log(\text{ADV})$ as in (4). For each portfolio allocation scheme, we sort stocks into five quintiles based on their assigned weights and compare the distribution of portfolio weights across the size quintiles to assess how heavily each size group is represented in the portfolio. The size composition of the top weight quintile is particularly informative, because it reveals the relative representation of small- versus large-cap stocks among the most heavily weighted positions in the portfolio.

Figure 1 presents heatmaps of the average proportion of stocks in each portfolio weight quintile that fall into each market value quintile, for the two portfolios we consider. Panel A shows the results for $\widehat{r}$ weighted portfolio, and Panel B corresponds to the portfolio with additional ADV weight. Each cell reports the average proportion of stocks in weight

quintile i that belongs to size quintile j, i.e., it illustrates the size distribution of stocks across portfolio weight quintiles.

Figure 1: Size Composition of Momentum Portfolio

Heatmaps of the average proportion of stocks in each portfolio weight quintile that fall into each market value quintile. The y-axis denotes portfolio weight quintiles (Q1 = lowest weight, Q5 = highest weight), the x-axis denotes size quintiles (Q1 = smallest firms, Q5 = largest firms), and each cell reports the average proportion of stocks in weight quintile i that belongs to size quintile j. Panel A shows the results for $\hat{r}$ weighted portfolio, while Panel B corresponds to the $\hat{r}\log(\text{ADV})$ weighted portfolio.



The last row of Panel A shows that, among the stocks receiving the highest portfolio weights (i.e., weight quintile 5), more than 40% are small-cap stocks (drawn from size quintiles 1 and 2). The average market value of these stocks is \$110M, and their average ADV is \$910K. In contrast, large-cap stocks from size quintile 5 (whose average market value is \$18B and average ADV is \$145M) constitute only 15% of the most heavily-weighted positions. Such allocations severely limit the portfolio's trading capacity, because the liquidity of small-cap stocks is insufficient to support large position sizes typical of institutional investors.

Incorporating the ADV weight into the portfolio construction process does not fully resolve the overallocation on small stocks. Panel B displays the results for the portfolio with additional capacity weight defined in (4). Although a scaling factor log(ADV) reduces exposure to small-cap names, the effect is partial. The proportion of small-cap stocks (quintile 1) among the highest-weighted positions decreases from 19% to 14%, while the share of large-cap stocks increases from 15% to 21%. However, small-cap stocks

continue to represent a substantial portion of the portfolio's most heavily weighted holdings. This highlights a persistent implementation challenge for this momentum strategy, because the limited tradable volume of small-cap securities constrains the scalability of such portfolios.

Furthermore, the return on a baseline portfolio value of \$1 is typically expressed as percentage terms (in basis points) to normalize performance relative to capital. While such normalization ensures consistency in comparison across time and strategies, it provides no information on the tradable volume or the achievable dollar return. Yet, trading capacity is a critical consideration when evaluating strategy profitability: the value of a 1% return on \$1 is different from that of a 1% return on \$1 million. Much of the literature on asset pricing anomalies focuses on the percentage return (i.e., how well one can trade), while largely overlooking the equally important question of how much one can trade.

Overall, capacity constraints make the weight allocations implied by these portfolio designs largely unimplementable in practice. Therefore, the metric Sharpe_H in (5) is a hypothetical Sharpe ratio rather than an achievable performance measure (hence the subscript "H"). The portfolio return in (4) remains a predictability metric, because it measures the covariance between predicted and realized excess returns and captures how well expected returns align with actual outcomes. However, this hypothetical Sharpe ratio provides no insight into whether such predictability can lead to scalable profits.

To assess the true economic contribution of the anomaly-based strategies, the evaluation measures need to incorporate the capacity constraints at a stock-level dollar allocation to determine how much one can trade.

Prior studies analyze the trade-off between trade size and the associated market impact costs to assess the break-even capacity at the portfolio level (Frazzini et al., 2012; Novy-Marx and Velikov, 2015). In contrast, our analysis takes a bottom-up approach. The framework focuses on tradable volume constrained solely by the capacity of individual stocks, independent of any trading costs incurred. This distinction is critical because our results show that capacity constraints alone can substantially limit the trade size of the achievable investment strategy, even before accounting for any trading costs. For completeness, we also incorporate trading costs to acknowledge their additional role in further reducing both tradable volume and realized returns.

In the following subsection, we focus on the practical capacity constraints present in

real-world trading environments and detail how to incorporate these into realistic trading strategies and in their profitability assessment.

Implementable Value and Dollar Return. Although the total number of a stock's outstanding shares determines an upper bound on the tradable volume of the stock, in reality, only a small fraction of these shares are available for trading on a given day. In our sample, the total market daily trading volume for individual stocks represents, on average, less than 1% of their outstanding shares.

A more practical reference of trade size is the stock's ADV, which reflects the latest trading activity in the market. In our sample, ADV is highly correlated with market value (approximately 80% on average), consistent with the well-established finding that larger companies tend to have more liquid and actively traded shares than smaller ones. Rather than assigning signal-based weights within a fixed portfolio budget, it is more realistic for traders to evaluate the tradable volume on a stock-by-stock basis.

We build our framework as follows: for a single investor, the feasible trade size of a stock is a percentage of its ADV. In a frictionless setting (i.e., assuming zero trading costs), a trader would optimally execute larger orders for stocks with higher absolute predicted returns, provided that the order size remains within the stock's tradable capacity. As such, the estimated trade size is scaled both by the strength of the predictive signal and by the stock's ADV. The dollar volume trade size for stock i at time t is

$$S_{i,t} = \min\left(\delta \, \mathrm{ADV}_{i,t} \, |\widehat{r}_{i,t}| \; \times 100 \, , \; \phi \, \mathrm{ADV}_{i,t} \, , \; \mathrm{Cap} \right) . \tag{6}$$

The first term, $\delta \, \mathrm{ADV}_{i,t} \, |\widehat{r}_{i,t}|$, estimates the target notional to trade, scaled jointly by signal strength $|\widehat{r}_{i,t}|$ and capacity proxy $\mathrm{ADV}_{i,t}$, and the scaling factor $\delta \in (0, 1)$ reflects the percentage of ADV the investor is willing to trade given the expected return is 1%. Alternatively, if we aggregate $S_{i,t}$ across all stocks and interpret the average of $|\widehat{r}_{i,t}|$ as a proxy for overall market signal strength, δ can be viewed as the fraction of the total market the investor is willing to engage with. This links the scaling factor δ directly to the investor's investment horizon and risk appetite.

The feasible participation rate $\phi \in (0, 1)$ imposes an upper limit on how much of a single stock's ADV can be traded. For example, Jegadeesh and Wu (2022) report that the closing auction volume averages around 10% of ADV, hence it is unlikely that a single

investor or institution will place an order matching the entire auction volume.

Cap (in USD) is an absolute limit on the trading amount per stock per day. This is particularly relevant for mega stocks (e.g., Apple, Meta), where a small fraction of ADV could result in excessively large trades in dollar terms. The Cap imposes a hard limit on the maximum dollar amount the investor is willing to buy or to sell in a single name.

The dollar volume $S_{i,t}$ ensures that trade sizes remain consistent with each stock's trade capacity limit and, more importantly, provides a direct measure of how much can be traded, both at the individual stock level and for the overall strategy.

It is well established in the literature that trading costs play a critical role in eroding the realized returns of anomalies. The standard approach to account for these costs is to subtract the estimated costs directly from the portfolio's gross returns (Novy-Marx and Velikov, 2015; DeMiguel et al., 2020; Chen and Velikov, 2023). However, in realistic trading scenarios, trading costs reduce gross profits and influence execution decisions. If a trader expects a stock's return to fall short of the associated costs, the trade is unlikely to be executed.[3]

To reflect this more realistic behavior, we adopt an ex-ante adjustment that filters out trades with insufficient expected return. The trade size after accounting for cost is

$$S_{i,t} = \begin{cases} \min\left(\, \delta \, \mathrm{ADV}_{i,t} \, |\widehat{r}_{i,t}| \, \times 100 \,,\ \phi \, \mathrm{ADV}_{i,t} \,,\ \mathrm{Cap} \,\right) \,, & \text{if } |\widehat{r}_{i,t}| > C_{i,t} \,, \\ 0 \,, & \text{otherwise} \,, \end{cases} \tag{7}$$

so the trade is executed only if the absolute predicted return exceeds the estimated total trading cost $C_{i,t}$. This mechanism ensures that trades with marginal or unprofitable signals are excluded ex-ante. The resulting dollar return, which we refer to as dollar profit and loss (PnL) to distinguish it from percentage return, is given by

$$\mathrm{PnL}_t \ = \ \sum_{i=1}^{N} \Big[\, S_{i,t} \, \mathrm{sign}(\widehat{r}_{i,t}) \, r_{i,t} \ - \ S_{i,t} \, C_{i,t} \Big] \,. \tag{8}$$

The first term on the right-hand side of (8) denotes the gross PnL of taking the position

[3]Some trades, such as hedging positions, may be executed at a loss yet still contribute positively to a broader strategy's overall risk-adjusted performance. Instead, our analysis focuses on directional strategies that take positions based on expected returns. These trades seek profits from predicted price movements rather than risk mitigation, thus, we apply the forward adjustment to reflect more realistic trading behavior.

$S_{i,t}\,\text{sign}(\hat{r}_{i,t})$ in stock i, and the second term is the trading costs scaled by order size for the stock. This provides an estimate of realized profitability with capacity constraints and trading costs.

As the final step, we normalize the PnL estimates by the corresponding total dollar volume traded and define the percentage return

$$\Psi_t = \frac{\text{PnL}_t}{\sum_i S_{i,t}} \times 100\% \,. \tag{9}$$

With these normalized returns, we compute the realized return t-statistic and annualized achievable Sharpe ratio

$$t_R = \frac{\overline{\Psi}}{s(\Psi)}\sqrt{T} \quad \text{and} \quad \text{Sharpe}_R = \frac{\overline{\Psi}}{s(\Psi)}\sqrt{252}\,, \tag{10}$$

where $\overline{\Psi} = \frac{1}{T}\sum_{t=1}^{T}\Psi_t$ is the sample mean of realized return, and $s(\Psi)$ is the sample standard deviation.

3 Tradable Volume of Anomalies

Next, we re-evaluate the performance of trading strategies that incorporate capacity constraints using our framework. Here, we set aside trading costs and focus on the impact of capacity limits; thus, all return estimates reported in this section are pre-cost, i.e., we set $C_{i,t} = 0$ in (7) and (8). This provides a direct measure of the tradable scale implied by an anomaly- or model-based strategy. First, we evaluate the performance of individual asset pricing anomalies, after which we evaluate predictive models that combine the predictors.

3.1 Single Anomaly Reassessment

We re-examine the predictive power of each anomaly as documented in the literature, and we evaluate the economic significance of these anomalies after accounting for real-world implementation limits. In particular, for each anomaly, we estimate a univariate regression model of market excess returns on the anomaly signal over a designated sample period. Then, we use the performance metrics introduced in the earlier section to evaluate the return predictions generated from the fitted model.

3.1.1 In-Sample and Out-of-Sample Analysis

To assess the stability and robustness of the predictive anomalies and the associated portfolio returns, we split our data into an IS period that aligns with the original publication's sample period and an OOS period that follows thereafter.

The literature has long debated the challenges of OOS replication and the risks of data mining in the discovery of return predictors (Jesen et al., 2023). Generally, one observes substantial declines in predictive R^2 and portfolio performance when models are tested outside the original sample, potentially due to overfitting, data-snooping biases and post-publication decay (Lo and MacKinlay, 1990; Campbell and Thompson, 2007; McLean and Pontiff, 2016).

For each single anomaly, across both sample periods, we track the evolution of parameter estimates and record the following metrics: R^2 score (3), hypothetical Sharpe_H (5), pre-cost dollar volume trade size S_t (6), pre-cost dollar return PnL_t (8), realized percentage return Ψ_t (9), t-statistics and realized Sharpe_R (10). We compare performance within and across the two sub-periods to isolate the effect of OOS performance deterioration, and to detect potential tradability issues despite high predictability in both periods, as would indicate a large drop from Sharpe_H to Sharpe_R.

3.1.2 Univariate Bayesian Regression

We specify the univariate linear model

$$r_{i,t} \;=\; X^k_{i,t-1}\,\beta^k \;+\; \eta_{i,t}\,, \quad i = 1, \ldots, N\,, \tag{11}$$

to capture the relationship between market excess returns and a given predictor. Here $X^k_{i,t-1} \in \mathbb{R}$ represents the exposure of stock i to predictor k at the end of time $t-1$, the regression coefficient β^k captures the sensitivity of returns to the predictor, and η_t is the unexplained component of return.

This formulation uses a single "pooled" regression across all stocks to estimate the sensitivity of stock returns to a given anomaly. Our primary focus is to assess the overall predictability of the anomaly for stock returns rather than to isolate effects for particular subsets of stocks. In other words, the coefficient β^k is common across stocks at a given time point, which is consistent with the general prediction error model in (2).

As discussed, we use R^2 (3) and hypothetical Sharpe_H (5) in predictive performance evaluation. For consistency in the analysis, we incorporate these weights into the regression training process. Similar to the approach for estimating time-varying market betas in (1), we use OBR to update β^k sequentially as new information becomes available.

In particular, let $\widehat{\beta}^k_{t-1}$ and Λ^k_{t-1} denote the prior mean and precision matrix (i.e., the inverse of the posterior variance) of the parameter β^k, respectively. With data from period $[t-1, t]$, the updates of the precision matrix and the posterior mean at time t are

$$\begin{aligned}
\Lambda^k_t &= \lambda\, \Lambda^k_{t-1} + {\boldsymbol{X}^k_{t-1}}^\top \boldsymbol{W}_{t-1}\, \boldsymbol{X}^k_{t-1}\,, \\
\widehat{\beta}^k_t &= \big(\Lambda^k_t\big)^{-1}\big(\Lambda^k_{t-1}\, \widehat{\beta}^k_{t-1} + {\boldsymbol{X}^k_{t-1}}^\top \boldsymbol{W}_{t-1}\, \boldsymbol{r}_t\big)\,,
\end{aligned}$$

where $\lambda \in (0,1)$ is the decay factor that downweights older observations, $\boldsymbol{X}^k_{t-1} \in \mathbb{R}^N$ is the vector of predictor k, $\boldsymbol{W}_{t-1} = \text{diag}(w_{1,t-1}, w_{2,t-1}, \ldots, w_{N,t-1})$ denotes the diagonal weighting matrix, with $w_{i,t-1}$ representing the weights assigned to stock i at time $t-1$, and $\boldsymbol{r}_t \in \mathbb{R}^N$ is the vector of market excess returns observed between $[t-1, t]$.

Finally, the OOS prediction from the anomaly for the market excess return in period $[t, t+1]$ is

$$\widehat{r}_{i,t+1} = X^k_{i,t}\, \widehat{\beta}^k_t\,,\ i = 1, 2, ..., N\,.$$

Of the 126 predictors, 21 are updated daily while the remaining 105 (from the CZ dataset) are updated every month. In the single anomaly reassessment, we align posterior updates with the frequency of each predictor, fitting the monthly predictors on monthly excess returns and the daily predictors on daily excess returns. For consistency in performance evaluation, we convert all dollar volume and return metrics to a daily frequency and annualize the resulting Sharpe ratios.

3.1.3 Result: Single Anomaly Performance

The full list of results for each individual anomaly is in Appendix E. Table 1 summarizes the average performance of trading single anomalies across two sample periods. Panel A reports the predictability metrics. In the IS period, 65% of single-anomaly strategies achieve hypothetical t-statistics exceeding the common benchmark of 1.96. This confirms the reproducibility of most anomalies documented in Chen and Zimmermann (2022). From IS to OOS, the R^2 decreases by an average of 52%, and the p-value of the one-

sided paired t-test (H_1: OOS R^2 < IS R^2) is less than 0.005. Similarly, the hypothetical Sharpe ratio declines by 39% on average, with a p-value of 2.0×10^{-11}. These results are consistent with the well-documented deterioration in anomaly performance when moving from IS to OOS evaluation.

Table 1: Average Performance of Anomalies

Average performance of single anomalies across IS and OOS periods. Panel A reports predictability metrics: the weighted R^2 in % (3), hypothetical Sharpe ratio Sharpe_H and the t-statistic t_H (5). Panel B presents the profitability metrics: dollar trade size S_t in \$K (6), realized dollar PnL_t in \$ (8), realized percentage return Ψ_t in % (9), realized Sharpe ratio Sharpe_R and the corresponding t-statistic of realized return t_R (10). The baseline capacity parameters are $\{\delta = 1\%, \phi = 5\%, \text{Cap} = \$1\text{M}\}$, and trading costs are set to zero, i.e., $C_{i,t} = 0$.

	IS	OOS
Panel A. Predictability		
Average R^2	0.02	−0.05
Average Sharpe_H	0.75	0.28
Number of anomalies with $t_H > 1.96$	82	50
Ratio of anomalies with $t_H > 1.96$	0.65	0.40
Panel B. Profitability under capacity constraints		
Average daily dollar volume traded S_t	240	3,000
Average daily PnL_t	710	2,100
Average daily realized return Ψ_t	0.15	0.06
Average Sharpe_R	0.32	0.13
Number of anomalies with $t_R > 1.96$	54	24
Ratio of anomalies with $t_R > 1.96$	0.43	0.19

Nevertheless, many anomalies retain a meaningful degree of predictability even in the OOS period. The average hypothetical Sharpe_H of the one-dollar portfolio remains above 0.5, and 40% of the anomalies have statistically significant hypothetical percentage return with $t_H > 1.96$.

Next, we examine the profitability of individual anomalies. Recall that under capacity constraints, both the pre-cost PnL and the percentage return of the trade are computed based on the estimated dollar volume traded, as defined in (6). This requires specifying the parameters that set the capacity limits. Table 1 Panel B presents results for the baseline specification where the investor can trade up to 1% of the ADV for each 1% of predicted return, subject to a hard cap of 5% of the ADV and of \$1 million per stock, i.e., set $\delta = 1\%$, $\phi = 5\%$, Cap = \$1M in (6). We use the realized percentage return series

of these constrained trades to compute t_R and Sharpe_R. All results for each individual anomaly is in Appendix E, Table 11.

Our results for the baseline parameters show that in both IS and OOS periods, the hypothetical Sharpe ratios are largely unachievable under realistic capacity limits. In the IS period, when anomaly predictability is the strongest, incorporating capacity constraints decreases the Sharpe ratio (from Sharpe_H to Sharpe_R) by an average of 40% (one-sided $p = 1.9 \times 10^{-12}$). During the OOS period, the decline is 22% due to capacity limits ($p = 6.1 \times 10^{-10}$), and the realized returns of only 24 anomalies are statistically significant. These results illustrate that before accounting for trading costs, capacity limits alone can substantially erode the profitability of return anomalies.

Recall that in our investment strategy the tradable volume is the minimum of three values: (i) the initial intended trade size, given by $\delta \, \text{ADV}_{i,t} \, |\widehat{r}_{i,t}| \times 100$; (ii) the maximum allowable trade size as a percentage of ADV, $\phi \, \text{ADV}_{i,t}$; and (iii) the absolute dollar Cap. One would expect different combinations of constraint parameters to lead to varying levels of tradable volume, and consequently, different realized dollar and percentage returns. To ensure that our findings are not driven by a specific parameter choice, we evaluate the robustness of our results for a range of capacity limit parameters, see Table 2.

Across all parameter choices in the table, from highly constrained to substantially relaxed, the realized returns of fewer than 30 anomalies are statistically significant (with $t_R > 1.96$). Even under a highly optimistic scenario (already largely unrealistic) that allows single-stock trades of up to 10% of their ADV and with a $10 million capital cap, the realized returns of only 20 anomalies are statistically significant. This highlights that the adverse effect of limited capacity on anomaly profitability is both persistent and robust to alternative specifications.

Given this robustness, we use the baseline parameters for consistency in all subsequent analysis. We repeat the experiments with different sets of parameters and obtain similar results, further reinforcing the robustness of the findings. To avoid redundancy, these additional results are not reported in the paper.

Next, we examine the relationship between predictability and profitability evaluation metrics. Figure 2 presents the scatter plots of the hypothetical Sharpe ratios Sharpe_H against the realized Sharpe ratios Sharpe_R for both IS (Panel A) and OOS (Panel B) periods. Each scatter point represents the performance of one anomaly.

Table 2: Performance of Anomalies Across Different Capacity Parameters

Number of anomalies with significant pre-cost realized percentage return (whose $t_R > 1.96$) for various capacity constraint parameters, as specified in (6). Recall that δ denotes the ADV % the investor is willing to trade per 1% of predicted return, ϕ sets the upper limit on the % of a single stock's ADV that can be traded, and Cap represents the absolute dollar cap on trade dollar volume.

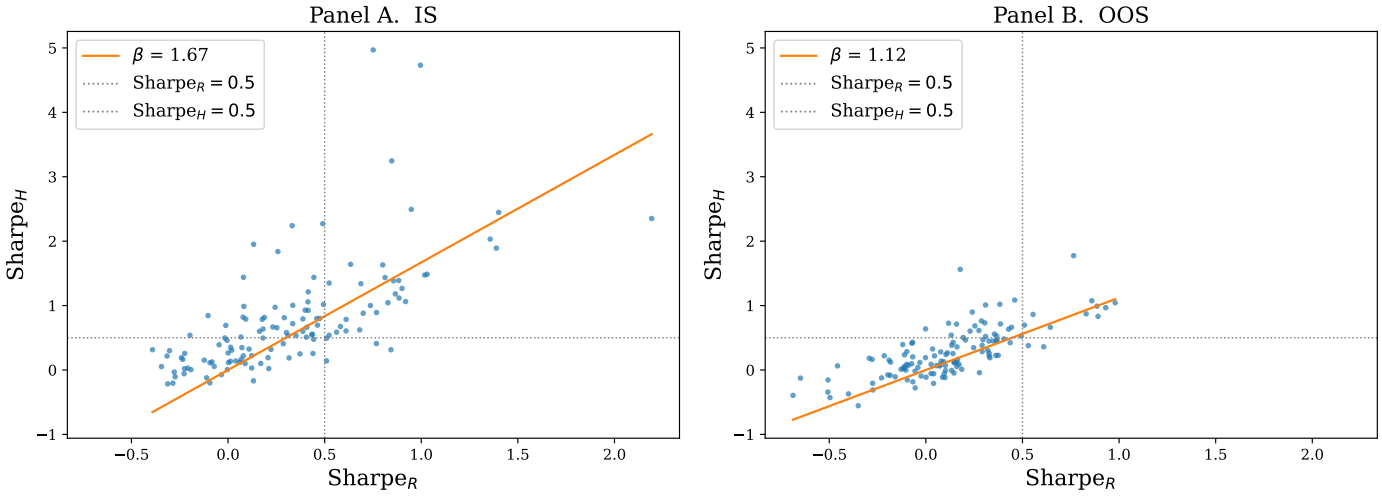
Cap ($)	$\phi = 0.5\%$	$\phi = 1.0\%$	$\phi = 5.0\%$	$\phi = 10.0\%$
Panel A. $\delta = 0.05\%$				
100K	18	18	18	18
1M	19	19	19	19
10M	19	17	17	17
Panel B. $\delta = 0.1\%$				
100K	24	24	24	24
1M	20	19	19	19
10M	20	19	18	18
Panel C. $\delta = 0.5\%$				
100K	24	24	25	25
1M	19	20	18	18
10M	20	21	19	19
Panel D. $\delta = 1.0\%$				
100K	22	23	23	23
1M	20	23	24	24
10M	23	22	20	19
Panel E. $\delta = 5.0\%$				
100K	26	27	28	28
1M	20	24	24	24
10M	23	23	19	20

In both panels of Figure 2, we observe a positive and statistically significant linear relationship between the predictability and profitability metrics. However, for almost all anomalies, Sharpe_R lies below Sharpe_H, i.e., regression slopes are greater than one. This pattern is consistent with the trend reported in Table 1.

We also include threshold lines at $\text{Sharpe}_H = 0.5$ and $\text{Sharpe}_R = 0.5$ as rough benchmarks for economic significance, which divide each panel of Figure 2 into four quadrants. In particular, many points appear in the upper-left quadrant, which indicates that the realized returns of a large number of single-anomaly strategies with strong predictive performance are not significant once capacity constraints are considered. In other words, hypothetical Sharpe ratios tend to overstate actual profitability. Even in the IS periods,

Figure 2: Single Anomaly Hypothetical Sharpe Against Realized Sharpe Ratio

Hypothetical Sharpe ratio, Sharpe_H, of each anomaly against its corresponding realized Sharpe ratio, Sharpe_R. Panel A displays IS results, while Panel B displays OOS results. Each scatter point represents the performance of one anomaly. Both panels include fitted zero-intercept regression lines. Benchmark lines are drawn at $\text{Sharpe}_H = 0.5$ and $\text{Sharpe}_R = 0.5$ as rough indicators for meaningful performance in terms of Sharpe ratios.



prior to any potential OOS deterioration or post-publication decay, the economic relevance of many anomalies is limited. Performance decline is more pronounced in the OOS period, where more than half of single-anomaly strategies with statistically significant Sharpe_H do not achieve significance in Sharpe_R. As a robustness check, we repeat the analysis on a restricted sample that includes only large-cap stocks (those in the top 20% market value quintile) and observe a similar decline in Sharpe ratios. Detailed results are reported in Appendix F.

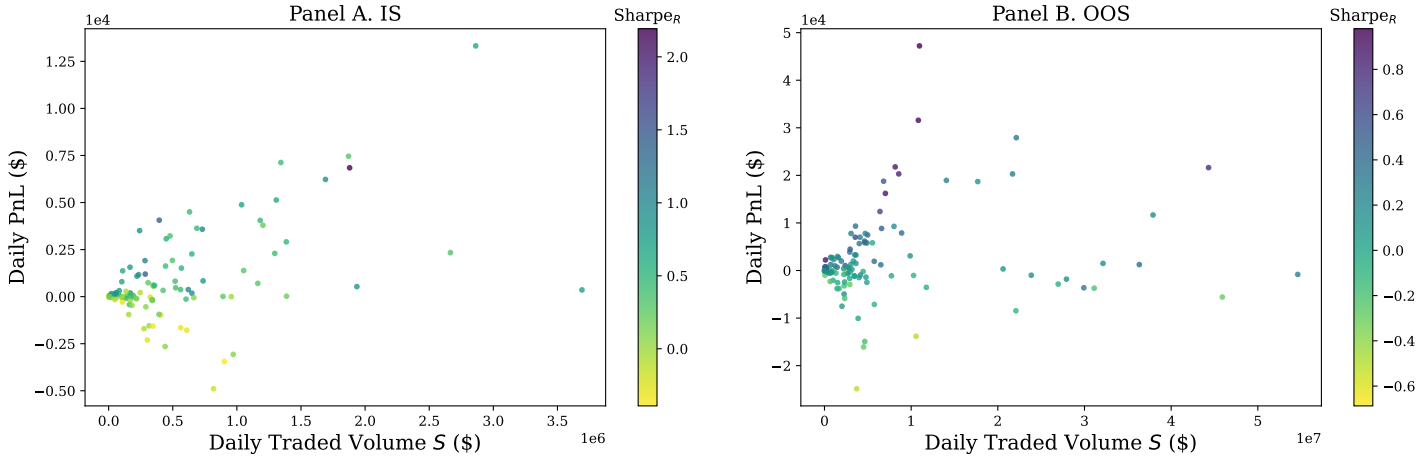
In addition, while higher predictability generally correlates with higher profitability, there are outliers with strong predictive performance but poor profitability. A notable example is one-day reversal, which records significant OOS hypothetical Sharpe_H of 1.56, yet has low profitability with realized Sharpe_R of 0.17. One-day reversal is not an isolated case; similar cases are in the upper left corner of both panels in Figure 2. These outliers are particularly misleading for investors because their seemingly strong predictive signals cannot be monetized into significant realized returns. This highlights the risk of relying solely on t_H and Sharpe_H as indicators for profitability: in some cases, these two metrics overestimate and misrepresent the performance of the anomaly-based investment strategies.

But does a high realized Sharpe guarantee economic significance? Not necessarily.

While the Sharpe ratio captures the risk-adjusted performance of percentage returns, it does not reflect the magnitude of trading volume or the scale of profits. Within our bottom-up capacity framework, trade sizes are capped as a fraction of ADV, which limits the trade size of strategies concentrated in illiquid stocks. As a result, an anomaly strategy may achieve a high Sharpe_R yet operate at modest trading volumes and generate limited realized dollar return. Such anomalies still offer meaningful predictive content, but the implementable value they accommodate varies widely. This distinction between statistical performance and tradable volume is important to keep in mind when evaluating anomaly-based strategies. To illustrate this disconnect, Figure 3 plots the pairs (PnL, S) pairs for the IS and OOS periods.

Figure 3: Single Anomaly PnL against Volume Traded

Scatter plots of the average daily dollar PnL, see (8), against the corresponding average daily dollar volume traded S, see (6). Each point represents one anomaly-based strategy. Panel A displays IS results, and Panel B displays OOS results. The color intensity of the scatter points reflects the anomaly's realized Sharpe_R, where darker points indicate higher realized Sharpe values. This visualizes the link among dollar volume traded, realized dollar PnL, and risk-adjusted percentage return.



Panels A and B of Figure 3 show that many scatter points cluster near the origin, with both traded volume S and PnL close to zero. Over 60% of single-anomaly strategies generate less than \$1,000 PnL during the OOS period before accounting for trading costs. This reflects that many anomalies operate on signals where implementable trade sizes are naturally constrained.

Figure 3 also reveals that several anomalies with relatively strong realized Sharpe_R (indicated by high color intensity) lie in regions with low tradable volume. Table 3 lists the top 15 anomalies ranked by realized Sharpe_R, and within this group, tradable volume and dollar profitability span a wide range. For example, the anomaly share turnover

volatility delivers a relatively high Sharpe$_R$ of 0.9, yet its average daily trading volume S is \$0.1M, with an average daily PnL of \$2.2K. Half of the top-performing anomalies in realized Sharpe generate pre-cost PnL of less than \$10K. The final "surviving" anomalies are mostly drawn from two economic categories: profitability and external financing. This pattern is consistent with the idea that profitability- and financing-based signals are more likely to operate in larger, more liquid firms, which may explain their stronger performance after accounting for capacity constraints.

Table 3: Top Performing Anomalies by Realized Sharpe Ratio

Ranked by OOS realized Sharpe$_R$, the top 15 anomalies that exhibit statistically significant pre-cost realized percentage returns after accounting for capacity constraints. For each anomaly, we present the hypothetical Sharpe$_H$, average daily PnL (\$K), average daily traded volume S (\$M), realized Sharpe$_R$, t-statistic, and realized percentage return Ψ (%). See Table 11, Appendix E for full list of the performance of anomaly-based investment strategies.

Anomaly	Sharpe$_H$	PnL	S	Sharpe$_R$	t_R	Ψ
Net external financing	1.0	47.2	11.0	1.0	4.8	0.4
Net equity financing	1.0	31.6	10.8	0.9	4.6	0.3
Cash-based operating profitability	0.8	16.2	7.0	0.9	2.8	0.3
Share turnover volatility	1.0	2.2	0.1	0.9	4.8	1.8
Operating profitability R&D adjusted	1.1	21.8	8.2	0.9	2.7	0.3
Gross profits / total assets	0.9	20.3	8.6	0.8	3.1	0.3
Five-day reversal	1.8	21.6	44.3	0.8	4.0	0.0
Operating leverage	0.7	12.4	6.4	0.6	2.6	0.2
Return seasonality years 6 to 10	0.4	7.0	3.5	0.6	2.9	0.2
Return on assets	0.4	4.5	3.0	0.5	2.3	0.2
Earnings forecast revisions	0.5	0.5	0.2	0.5	2.8	0.3
Earnings announcement return	1.1	8.8	6.6	0.5	2.6	0.2
Share issuance (1 year)	0.7	0.2	0.1	0.4	2.0	0.2
Analyst earnings per share	0.4	0.8	0.0	0.4	2.0	0.3
Net debt financing	0.6	6.2	4.6	0.4	2.0	0.3

These findings do not suggest that the anomalies with lower tradable volume are useless. Rather, they highlight that the economic impact of anomalies varies widely when trading capacity is taken into account. In practice, anomalies with different implementable scales may offer distinct strategic value to investors with different capital bases. Tradable volume and dollar returns therefore represent a complementary and practically important dimension of profitability. We revisit how investors of different fund sizes may make use of signals with varying scalability in Section 5.

Additionally, the hypothetical Sharpe$_H$ for half of the anomalies in Table 3 is below 1.0, which is generally not considered impressive in terms of predictability performance. This aligns with our earlier observation of a common mismatch between predictability and profitability measures. As a next step, we investigate which groups of stocks drive high predictability or profitability, and study the underlying drivers of the frequent divergence between these two characteristics.

To better understand the gap between predictability and profitability results, we first examine two representative anomaly-based strategies. One-day reversal (Ret1d), which records significant Sharpe$_H$, yet yields poor Sharpe$_R$; and net external financing (XFIN), with the highest Sharpe$_R$ and dollar return, despite having a lower Sharpe$_H$. We break down the performance of these two strategies by stock size, splitting the stock universe into five quintiles (quintile 1 represents the smallest stocks and quintile 5 the largest) and present the results in Table 4.

Table 4: Performance by Size Quintiles: Ret1d and XFIN

Performance metrics of two anomalies: One-day reversal (Ret1d) and net external financing (XFIN) across size-sorted quintiles. Stocks are sorted into five quintiles based on market value, where quintile 1 represents the smallest stocks and quintile 5 the largest. Panel A reports the performance metrics for Ret1d, and Panel B presents the results for XFIN. For each anomaly, we present the hypothetical Sharpe$_H$, t_H, average daily PnL (\$K), average daily traded volume S (\$M), realized Sharpe$_R$, and t_R.

Market Value Quintile	Sharpe$_H$	t_H	PnL	S	Sharpe$_R$	t_R
Panel A. Ret1d						
1	2.9	15.1	0.1	0.1	0.2	1.0
2	0.9	4.9	0.6	0.5	0.2	0.9
3	0.1	0.7	0.6	1.5	0.0	0.2
4	−0.3	−1.7	1.7	5.0	−0.1	−0.3
5	0.0	0.1	8.9	30.8	0.2	0.8
Panel B. XFIN						
1	1.7	8.3	0.4	0.0	1.0	4.9
2	1.1	5.4	2.7	0.2	1.2	5.7
3	0.4	1.7	5.3	0.6	0.7	3.6
4	0.4	2.0	7.9	1.7	0.3	1.5
5	0.9	4.4	30.9	11.0	1.1	5.2

Table 4 shows that the high overall Sharpe$_H$ of Ret1d is mainly driven by quintiles 1. However, due to capacity constraints, the tradable volume S for these small-cap stocks remains around \$1,000K per day. Meanwhile, Ret1d performs poorly on large-cap stocks, yet these are the stocks for which sufficiently large trade sizes are feasible. As

a result, the small profits generated from small-cap trades are outweighed by the losses incurred on larger trades in less predictive large-cap stocks. This underscores the critical limitation of the hypothetical Sharpe ratio: although it incorporates ADV weights (akin to value-weighting in the literature), it does not capture the true economic significance of an anomaly when scaled to realistic trading constraints. XFIN, on the other hand, maintains relatively strong predictability in large-cap stocks and can execute profitable trades on a larger scale, despite its lower predictability in small-cap stocks.

Figure 4 presents the distribution of performance metrics by size quintiles across all single-anomaly strategies. Panel A shows the box plots of Sharpe_H. In general, small-cap stocks tend to exhibit higher hypothetical Sharpe ratios than those of larger-cap stocks. This is consistent with the common understanding that small-cap stock markets are more inefficient and thus more predictable (Avramov, 2002). Panel B shows the distribution of the realized Sharpe ratios where the advantage of small-cap predictability becomes less pronounced after accounting for capacity constraints.

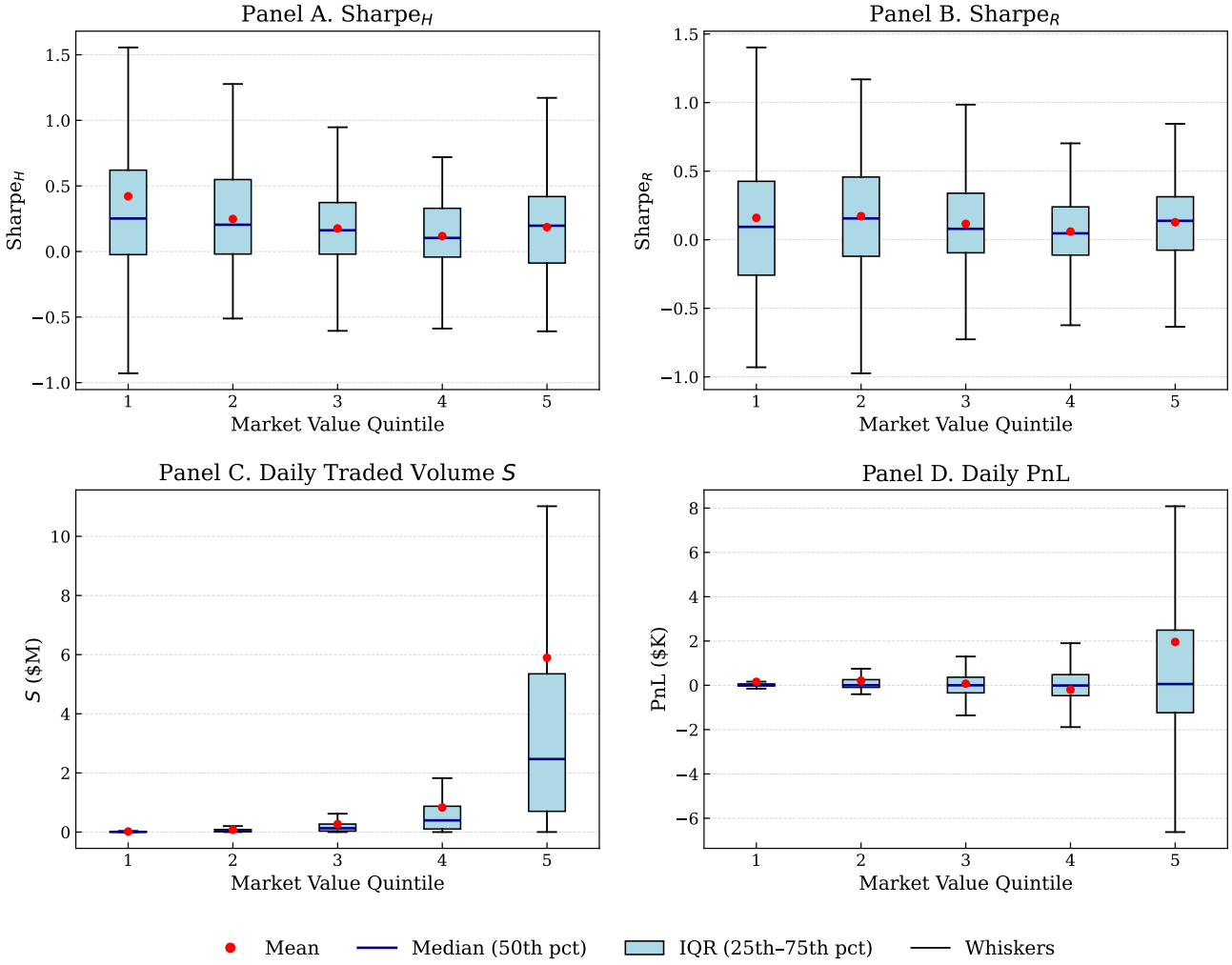
The key insight is in Panels C and D. The tradable volume and dollar return are heavily concentrated in large-cap stocks (particularly those in quintile 5). Smaller stocks (quintiles 1 and 2) contribute very little to total PnL despite their potentially higher predictability. Capacity constraints substantially limit the feasible trading volume among small-cap stocks, reducing their impact on realized economic gains. Consequently, anomalies such as Ret1d, which exhibit strong predictive power primarily in illiquid small-cap stocks, do not generate meaningful realized profits and are less tradable in practice. This illustrates how the mismatch between return predictability and trading capacity can lead to a pronounced disconnect between statistical predictability and real-world profitability.

3.2 Combination of Anomalies

A substantial body of literature generally reports that the R^2 score and portfolio Sharpe ratios of models combining multiple anomalies tend to be higher than those of univariate regression models based on single predictors (Haugen and Baker, 1996; Gu et al., 2020). Nonetheless, it remains largely unexplored whether these improvements in predictive performance generate greater implementable values and realized return after considering capacity constraints. Below, we extend our analysis to evaluate the predictive power of models that incorporate multiple anomalies, and examine whether their signals are

Figure 4: Distribution of the Performance Metrics of Anomalies by Size Quintiles

Box plots of the performance metrics for all single anomalies during the OOS period, grouped by quintiles of market value. Stocks are sorted into five quintiles based on market value, where quintile 1 represents the smallest stocks and quintile 5 the largest. Panels A reports the distribution of the predictability measure Sharpe_H across all the anomalies, while Panels B, C, and D show the profitability metrics Sharpe_R, daily traded volume S, and dollar PnL after accounting for capacity constraints respectively.



implementable with meaningful trade size in realistic market conditions.

3.2.1 Data and Model Tuning

We combine the anomalies assessed in the previous section. The monthly anomaly characteristics are expanded to a daily frequency by forward filling each feature with its most recent monthly data. The sample period for the combined dataset spans from January 1993 to December 2023. To mitigate IS bias when combining anomalies, each predictor is incorporated into the model after the conclusion of its original IS period. Specifically, model features are updated at the beginning of each calendar year to include any new

anomalies that entered their OOS phase.

This study employs a selection of widely-used models for stock return prediction, including both linear and non-linear approaches. We do not aim to develop a model architecture that maximizes predictive accuracy. Instead, our objective is to assess whether improvements in predictability, when achieved, can lead to tangible profits.

The linear models include multivariate Bayesian regression and penalized Bayesian regression. For non-linear modeling, we adopt the boosted regression tree (BRT) framework, which captures both non-linear effects and interactions among predictors (Gu et al., 2020). In addition, we evaluate two ensemble methods that combine forecasts from individual models to leverage complementary information: simple averaging and multivariate Bayesian model averaging. Detailed descriptions of the model algorithms are in Appendix A.2.

We implement a rolling window approach for sample splitting and hyperparameter tuning in the machine learning models, see Appendix B for details.

3.2.2 Results: Combined Anomalies

Table 5 presents the performance metrics for all predictive models introduced earlier. Panel A presents the predictive metrics, and Panel B shows the profitability metrics that account for capacity limits for the baseline parameters.

The predictive performance of all models is significant, with most weighted R^2 scores above 0.1% and the model hypothetical Sharpe_H exceeding 3, which is a substantial improvement over individual anomaly models. Similar to the single-anomaly case, model performance declines notably after accounting for limitations in tradable volume, as shown by the hypothetical Sharpe_H and the realized Sharpe_R. However, encouragingly, the tradable volume increases significantly across the models when compared to strategies based on single anomaly. All pre-cost realized returns are significantly positive, with $t_R > 1.96$ and most Sharpe_R above 1. This suggests that the predictive signals generated by models combining multiple anomalies retain a certain degree of profitability. Furthermore, the majority of these strategies generate an average daily pre-cost PnL above \$25,000.

Among the predictive models, the R^2, realized percentage return Ψ, and realized Sharpe_R of the BRT models are generally higher than those of the linear regressor. This

Table 5: Performance of Predictive Models

Summary statistics of the performance for predictive models that combine anomalies. The models are: multivariate Bayesian regression (MBR), penalized multivariate Bayesian regression (PMBR), XGBoost, LightGBM, CatBoost, ensemble model using simple averaging (Agg_Avg) and ensemble model using penalized regression (Agg_PMBR). Panel A presents the results for predictability metrics: R^2 (%) (3), Sharpe_H and t_H (5). Panel B records the results for pre-cost profitability metrics: S_t (\$M) (6), PnL_t (\$K) (8), Ψ_t (%) (9), Sharpe_R and t_R (10) computed with the baseline parameters $\{\delta = 1\%, \phi = 5\%, \text{Cap} = \$1\text{M}\}$. The best-performing model for each evaluation metric is in bold.

	MBR	PMBR	XGBoost	LightGBM	CatBoost	Agg_Avg	Agg_PMBR
Panel A. Predictability							
R^2	0.07	0.04	0.14	0.20	0.31	0.36	**0.36**
Sharpe_H	3.41	2.90	5.54	6.36	**7.21**	5.73	6.82
t_C	18.94	15.36	29.26	33.58	**38.08**	30.28	35.41
Panel B. Profitability with capacity constraints							
S_t	96.97	93.14	82.95	**102.27**	75.60	72.28	74.43
PnL_t	29.68	28.61	28.02	24.99	26.8	28.72	**45.82**
Ψ_t	0.03	0.04	0.04	0.04	0.04	**0.05**	0.05
Sharpe_R	0.88	0.93	1.03	1.03	1.04	**1.39**	1.39
t_R	4.91	4.92	5.46	5.43	5.52	**7.32**	7.22

finding, consistent with the results in Gu et al. (2020), shows that tree-based models deliver superior predictive performance relative to simple linear regressions. Moreover, the two ensemble models demonstrate the strongest performance across both predictability and profitability metrics.

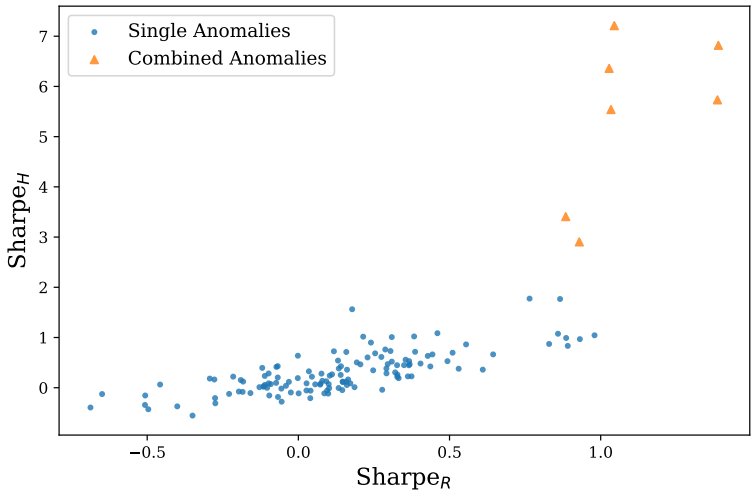
Figure 2 plots the predictability and profitability evaluation metrics. In addition to the performance of single-anomaly strategies (represented by blue dots), Figure 5 includes the performance of models that combine anomalies (represented by orange triangles). Combining the anomalies substantially enhances predictive performance with all models demonstrating higher hypothetical Sharpe_H than any single-anomaly strategy. However, the corresponding improvement in realized Sharpe_R is modest after accounting for capacity constraints.

Recall that in the single-anomaly reassessment, we emphasized that absolute trading volume and dollar profits are as important as percentage returns when evaluating the profitability of a strategy. The plots of the realized Sharpe ratios and dollar PnL against the trading volume in Figure 6 reinforces this message

Models that combine multiple anomalies (represented by orange triangles) predominantly cluster in the upper-right quadrant of both panels in Figure 6. This pattern

Figure 5: Hypothetical Sharpe Ratios Against Realized Sharpe Ratios

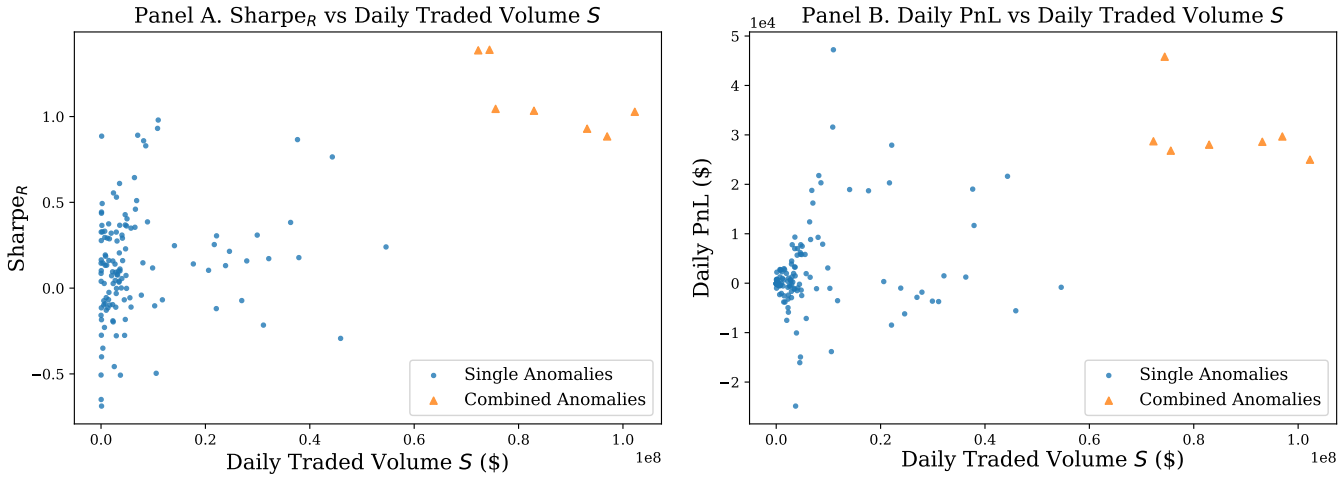
Hypothetical Sharpe ratio, Sharpe_H, against the corresponding realized pre-cost Sharpe ratio, Sharpe_R after accounting for capacity constraints. Each scatter point corresponds to a distinct anomaly- or model-based trading strategy. The blue dots represent the single anomalies, and the orange triangles represents the predictive models that combine multiple anomalies.



suggests that combining signals improves risk-adjusted returns and increases the scalability of the strategies.

Figure 6: Realized Sharpe and Dollar PnL Against Traded Volume

Panel A plots the realized Sharpe ratio, Sharpe_R, against the corresponding average daily traded volume S for each individual anomaly and for predictive models that combine multiple anomalies. Panel B plots the average daily dollar PnL against the trading volume. Each scatter point corresponds to a distinct anomaly- (blue dots) or model-based (orange triangles) trading strategy.



However, the increase in eligible trading volume does not necessarily generate a proportional increase in dollar PnL, see Panel B in Figure 6. While the predictive models exhibit higher trading volumes relative to all single-anomaly strategies, the corresponding gains in dollar PnL remain modest compared to the top-performing anomalies.

The finding highlights that improving predictive performance, while valuable, should not be viewed as the sole objective when developing or evaluating return forecasting models. A critical step is to assess if these improvements generate meaningful gains in scalable trading volume and dollar PnL. As the results suggest, gains in profitability metrics are often modest relative to the improvements observed in predictive accuracy. In many cases, enhanced predictability does not fully carry over to tangible, scalable profits, underscoring the importance of evaluating models both by their statistical performance and their practical implementability.

To understand better how each performance metric improves when anomalies are combined, we examine results across different size quintiles. Table 6 presents two illustrative examples: anomaly XFIN, which performs relatively well in terms of both predictive power and pre-cost profitability among single predictors; and model Agg_PMBR, identified earlier as one of the best-performing combination strategies. The table shows that moving from XFIN to Agg_PMBR, the predictive measure Sharpe_H improves in all quintile groups, where the largest increase is among small-cap stocks. The corresponding increase in post-cost realized Sharpe is more modest but follows a similar trend. In particular, the average daily PnL for the stocks in quintile 1 rises from \$0.4K to \$1.7K. Despite this significant percentage increase, the absolute scale of profits in small-cap stocks remains constrained by capacity limits. The additional gains from small stocks are small compared to the large PnL generated from large-cap stocks in quintile 5, where available trading volume is much higher. This reinforces our earlier observation: while combining anomalies greatly enhances predictability, it yields limited improvements in profitability.

For all predictive models, Figure 7 presents the distribution of performance metrics across size quintiles. Consistent with the results for Agg_PMBR, Panel A shows that all models exhibit strong predictive performance for stocks in the smallest size group (quintile 1), with performance generally declining as the market capitalization of stocks increases. A similar pattern is observed with the realized Sharpe ratios, see Panel B. However, Panels C and D show that the eligible trading volume and the dollar return generated from trading are heavily concentrated in large-cap stocks in quintile 5, despite their relatively low levels of predictability. These distributional patterns are similar to those documented for single-anomaly strategies, see Figure 4.

Table 6: Performance by Size Quintiles: XFIN and Agg_PMBR

Performance metrics of anomaly, XFIN, and the ensemble model Agg_PMBR across size-sorted quintiles. Stocks are sorted into five quintiles based on market value, with quintile 1 representing the smallest stocks and quintile 5 the largest. Panel A reports the performance metrics for XFIN, and Panel B presents the results for Agg_PMBR. For each strategy, we present the hypothetical Sharpe$_H$, t_H, average daily PnL ($K), average daily traded volume S ($M), realized Sharpe$_R$, and t_R.

Market Value Quintile	Sharpe$_H$	t_H	PnL	S	Sharpe$_R$	t_R
Panel A. XFIN						
1	1.7	8.3	0.4	0.0	1.0	4.9
2	1.1	5.4	2.7	0.2	1.2	5.7
3	0.4	1.7	5.3	0.6	0.7	3.6
4	0.4	2.0	7.9	1.7	0.3	1.5
5	0.9	4.4	30.9	11.0	1.1	5.2
Panel B. Agg_PMBR						
1	6.7	34.7	1.7	0.3	2.2	11.4
2	5.1	26.4	2.9	1.2	1.7	9.0
3	3.5	18.3	5.4	3.7	1.8	9.7
4	2.5	12.9	9.0	10.7	1.4	7.4
5	1.7	9.1	29.1	58.5	1.0	5.2

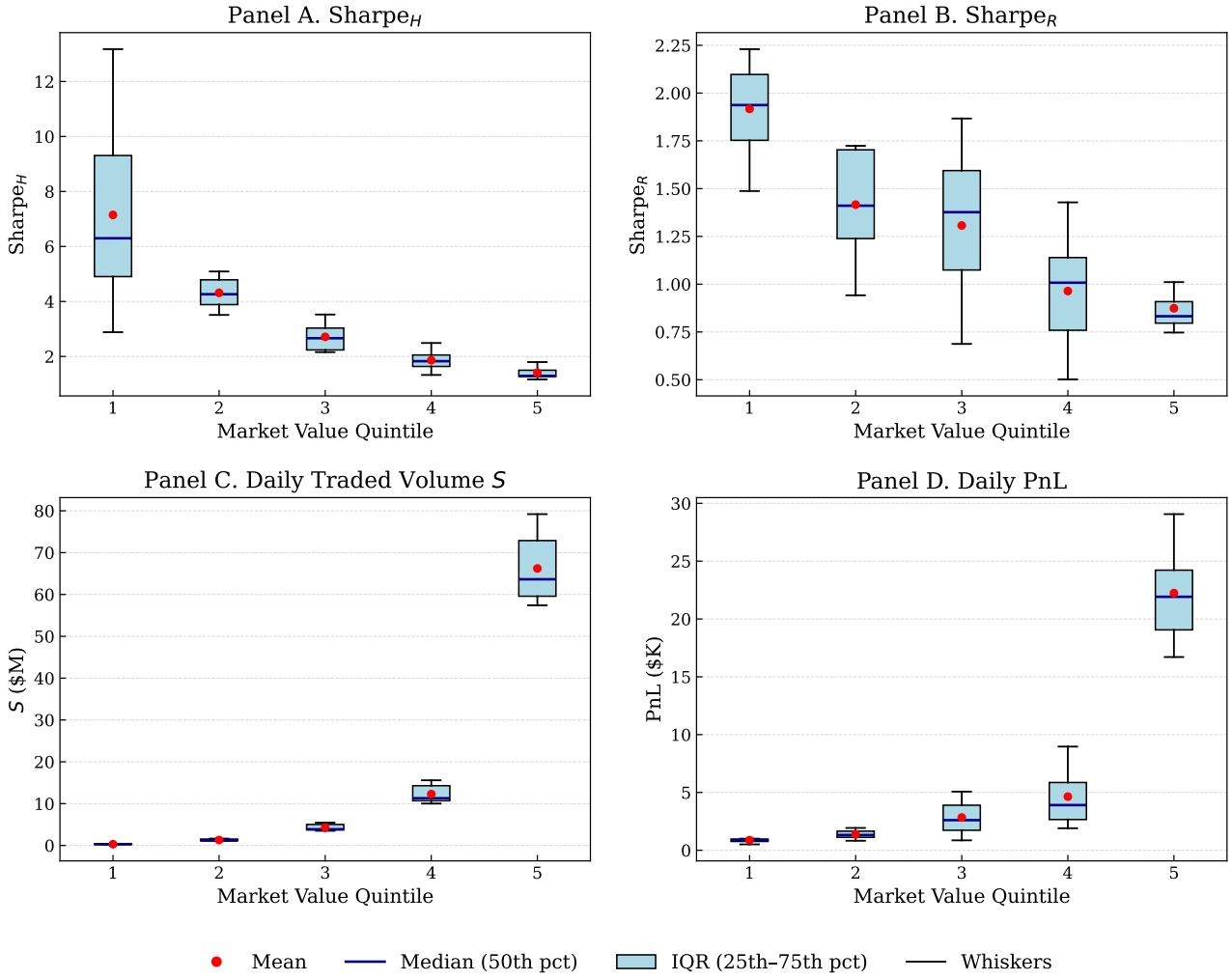
Our findings are consistent with those of prior studies that document improved return forecasts from various predictive models. However, the improvement in profitability is modest and far less optimistic than suggested by hypothetical Sharpe ratios. Most of the gains in predictability are concentrated in small-cap stocks. Yet, once capacity constraints are accounted for, these improvements do not generate substantial increases in absolute profits. In contrast, predictive gains for large-cap stocks are much more limited, which constrains the potential for meaningful improvements in economic significance.

4 Estimates of Trading Costs

Trading costs erode gross returns. In the previous section, our framework showed that in the absence of trading costs, capacity constraints alone can substantially limit tradable volumes and thereby cap the absolute profitability of a strategy. For completeness, we now include trading costs and examine their additional impact on strategy performance.

Figure 7: Distribution of the Performance Metrics of Models by Size Quintiles

Box plots of performance metrics for all predictive models that combine the anomalies during the OOS period. Stocks are sorted into five quintiles based on market value, where quintile 1 represents the smallest stocks and quintile 5 the largest. Panel A reports the distribution of the predictability measure Sharpe_H across all the anomalies, while Panels B, C, and D show the pre-cost profitability metrics Sharpe_R, daily S, and PnL respectively.



4.1 Trading Cost Estimation

In the existing literature, trading cost estimations are typically classified into two broad categories: the cost of crossing the bid-ask spread and the cost incurred by market price impact. In this paper, we adopt both measures to assess trading costs and ensure the robustness of our results.

The half-spread is widely used in the literature as a measure of transaction costs during regular trading hours (Box et al., 2021; Chen and Velikov, 2023). The intuition is that buyers are assumed to pay half the spread above the prevailing midpoint, while sellers receive half the spread below it. Following Chen and Velikov (2023), we construct low-frequency daily proxies for the half-spread, see details in Appendix C.1.

The half-spread estimate may substantially overstate daily realized return penalties. Low-frequency proxies are upward biased by approximately 25–50 basis points in the post-2005 period, see Chen and Velikov (2023). Moreover, in practice, traders often rely on passive orders rather than aggressively crossing the spread, which significantly reduces spread-related costs.

Instead, Frazzini et al. (2018) therefore argues that price-impact cost constitutes the dominant component of total transaction costs. Price impact captures changes in an asset's market price induced by trade execution, particularly in large volumes. Large orders can exhaust available order book supply, driving prices upward for purchases or downward for sales. These costs are especially pronounced for illiquid securities or sizable trades. We adopt the price impact model of Frazzini et al. (2018), see Appendix C.2 for details.

It is also worth noting that additional frictions, such as stock-borrowing fees associated with short-selling and various administrative or brokerage fees, could further erode the profitability of anomaly strategies. These costs are not modeled here but would only reinforce the downward pressure on returns documented in our analysis.

4.2 Results

Recall that our proposed framework incorporates trading costs ex-ante, so only trades whose expected returns exceed the cost estimates (as specified in (7)) are executed. The post-cost dollar PnL is computed based on the actual dollar trade sizes, see (8). Table 7 reports the average results of the profitability evaluation under zero cost, price impact cost, and half-spread cost estimates.

Panel A of Table 7 summarizes the average performance of single-anomaly strategies in the OOS period. All profitability measures decline once trading costs are included, with realized dollar PnL, percentage return, and Sharpe ratio more heavily penalized under half-spread costs compared to those that account for the price-impact costs.

On average, the realized Sharpe_R of single-anomaly strategies decreases by 16.1% from zero-cost to the price-impact setting (p-value $= 2.8 \times 10^{-6}$), and by 33.6% under half-spread setting (p-value $= 5.3 \times 10^{-11}$).

Recall from Table 1 that the number of anomalies with statistically significant realized percentage returns declines from 50 to 24 (as also shown under the "zero cost" column

Table 7: Average Profitability Performance under Different Costs

Average profitability performance for single anomalies during OOS periods (Panel A) and predictive models that combine anomalies (Panel B) for three cost specifications: zero cost, price impact cost based on Frazzini et al. (2018), and half-spread cost following Chen and Velikov (2023). The profitability performance metrics include dollar trade size S_t (\$M) (7), realized dollar PnL_t (\$K) (8), realized percentage return Ψ_t (%) (9), realized Sharpe ratio Sharpe_R and the corresponding t-statistic of realized return t_R (10). All metrics are computed with the baseline capacity parameters $\delta = 1\%$, $\phi = 5\%$, Cap = \$1M. The last row of each panel shows the fraction of trades that are executed, i.e., the predicted returns were higher than the cost threshold, i.e., $|\widehat{r}_{i,t}| > C_{i,t}$ and $V_{i,t} > 0$ in (7).

	Zero Cost	Price Impact	Half Spread
Panel A. Single Anomalies (OOS)			
Average daily dollar volume traded S_t	3.01	3.00	2.50
Average daily PnL_t	2.10	1.03	−0.15
Average daily realized return Ψ_t	0.06	0.05	−0.04
Average Sharpe_R	0.13	0.08	−0.00
Number of anomalies with $t_R > 1.96$	24	18	9
Ratio of anomalies with $t_R > 1.96$	0.19	0.14	0.07
Number of anomalies with $\text{Sharpe}_R > 0.5$	12	9	7
Ratio of anomalies with $\text{Sharpe}_R > 0.5$	0.10	0.07	0.06
Ratio of trade executed ($\|\widehat{r}_{i,t}\| > C_{i,t}$)	1.00	0.69	0.21
Panel B. Models Combining Anomalies			
Average daily dollar volume traded S_t	86.10	82.85	78.01
Average daily PnL_t	31.95	14.34	1.31
Average daily realized return Ψ_t	0.04	0.03	0.01
Average Sharpe_R	1.14	0.66	0.16
Number of models with $t_R > 1.96$	7	7	2
Ratio of models with $t_R > 1.96$	1.00	1.00	0.29
Number of models with $\text{Sharpe}_R > 0.5$	7	5	1
Ratio of models with $\text{Sharpe}_R > 0.5$	1.00	0.71	0.14
Ratio of trade executed ($\|\widehat{r}_{i,t}\| > C_{i,t}$)	1.00	0.98	0.44

in Table 7) when one accounts for capacity constraints. Including trading costs further reduces this number to 18 under price-impact costs and 9 under half-spread costs.

The average dollar trade size $S_{i,t}$ also declines when trading costs are considered, because a trade is only executed if the absolute expected raw return exceeds the cost estimate, i.e., $|\widehat{r}_{i,t}| > C_{i,t}$, see (7). The last row of Panel A in Table 7 reports the fraction of trades that meet this condition and thus have non-zero dollar volume. Accounting for price-impact costs, approximately 69% of the trades are executed, whereas under half-spread costs, only 21% are executed. For many anomalies, fewer than 5% of trades

generate expected raw returns that exceed the half-spread cost estimate.

This offers an alternative explanation for why much of the literature concludes that trading costs almost completely eliminate realized returns. Beyond the risk of over-penalization, studies that directly subtract total trading costs from portfolio percentage returns may implicitly include many trades with negative expected post-cost returns, trades that, in practice, would not have been executed in the first place.

The decline in profitability is further illustrated in Figure 8, which presents box plots of performance metrics for all single-anomaly strategies in different cost settings. Both Sharpe_R and realized dollar PnL exhibit a clear downward shift once trading costs are included. The distribution of tradable volume is largely unchanged after considering price-impact costs, because most trades continue to be executed. However, tradable volume declines markedly after imposing half-spread costs because many trades are screened out by the higher cost estimates.

Figure 8: Performance Distribution of Anomalies under Different Costs

Box plots of performance for all single anomalies that account for zero cost, price-impact costs and half-spread costs. Panel A reports the distribution of realized Sharpe ratios Sharpe_R (10), Panel B presents the distribution of realized dollar trade size S (7), and Panel C plots the distribution of daily realized dollar PnL (8).

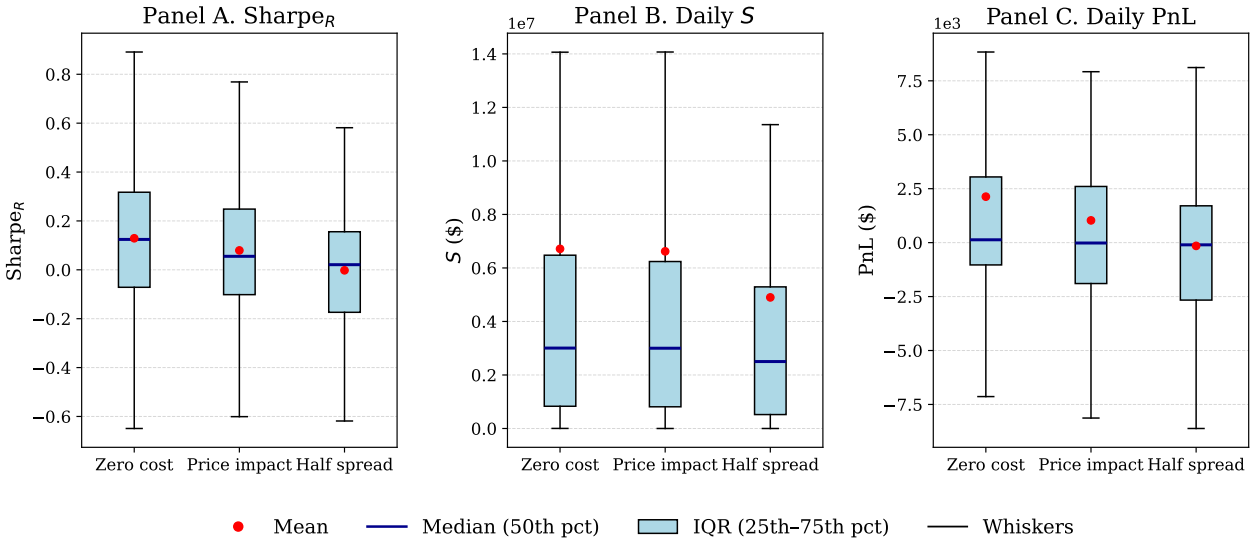
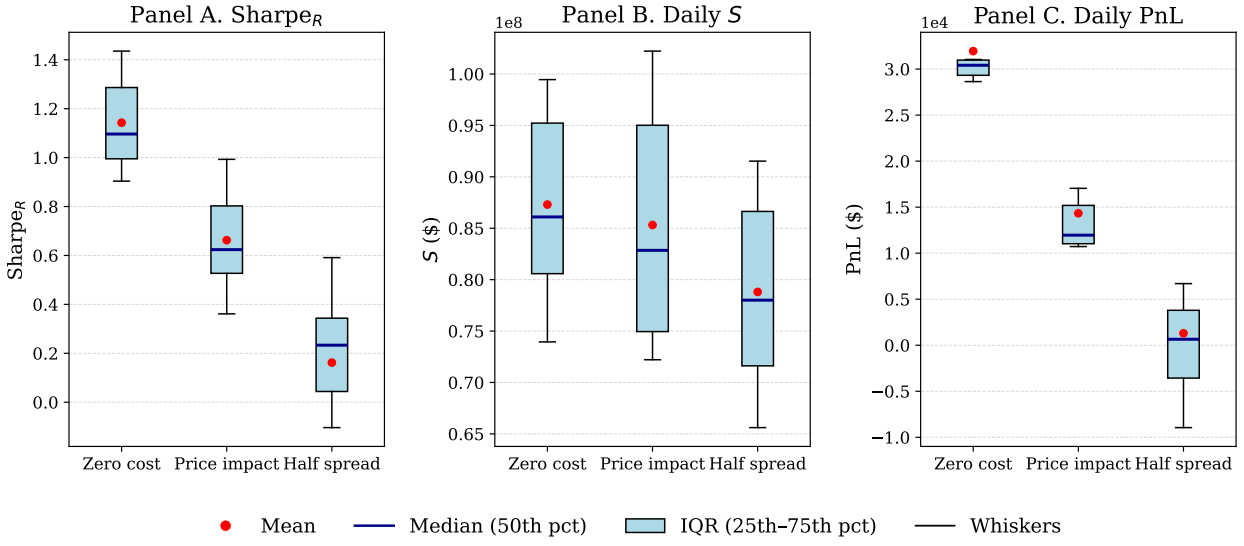


Table 7, Panel B presents the results after considering trading costs for predictive models that combine the anomalies, and the distributions of the evaluation metrics are in Figure 9. Similar to the trends observed among single-anomaly strategies, trading costs further erode the realized profits and limit the tradable volume of the associated strategies. The average realized Sharpe ratio Sharpe_R declines from 1.14 under zero cost setting to 0.66 under price-impact costs, while the average realized dollar PnL falls

sharply from \$32K to \$14K. The decline trends in both percentage and absolute returns are clearly visible in Panels A and C of Figure 9.

Figure 9: Performance Distribution of Models under Different Costs

Box plots of performance for all predictive models that combine the anomalies accounting for zero cost, price-impact costs, and half-spread costs. Panel A reports the distribution of realized Sharpe ratios Sharpe_R (10), Panel B presents the distribution of realized dollar trade size S (7), and Panel C plots the distribution of daily realized dollar PnL (8).



Models that combine multiple anomalies continue to outperform individual anomalies after accounting for trading costs. With the exception of the two linear regression models, the percentage returns under price-impact costs are statistically significant in all other predictive models. Notably, the two ensemble models once again exhibit the strongest performance after imposing both cost estimates.

One of the key advantages of combining anomalies is the increase in tradable volume. Models that aggregate multiple signals tend to produce more accurate and stronger predictions, yielding larger values of $|\widehat{r}_{i,t}|$. This serves two purposes: first, it increases the intended trade size in (6); and second, it increases the likelihood that expected raw returns are higher than estimated trading costs. When accounting for price-impact costs, 98% of trades are, on average, executed for predictive models (see Panel B, Table 7). Although this fraction drops to 44% when accounting for half-spread costs, it nevertheless represents a substantial improvement relative to the single-anomaly case.

Incorporating trading costs further erodes the performance of anomaly-based strategies. However, the impact of capacity constraints remains significant and is not overshadowed by trading costs. Both Table 1 and Table 7 show that only 24 anomalies retain statistically significant realized percentage returns despite the zero-cost assumption. In

addition, both cost estimates are in percentage terms (basis points), and thus the tradable volume estimate is still essential when computing dollar PnL from post-cost returns, an equally important dimension of profitability assessment.

It is also crucial to emphasize that trading costs should be viewed in addition to capacity constraints, rather than in isolation. Subtracting estimated costs from portfolio returns without accounting for the implementable trading scale could over or under estimate both the portfolio's performance and its sensitivity to trading constraints.

5 Hedge Fund Size and Performance

When a fund's capital base exceeds the tradable capacity of a particular stock, it must diversify into a broader set of securities, even if that stock offers higher expected returns.

In the final section of this paper, we present a case study that monitors the performance of a hypothetical fund for various levels of investment capital. This analysis illustrates that signals with different scalability offer distinct strategic value to investors depending on their capital base.

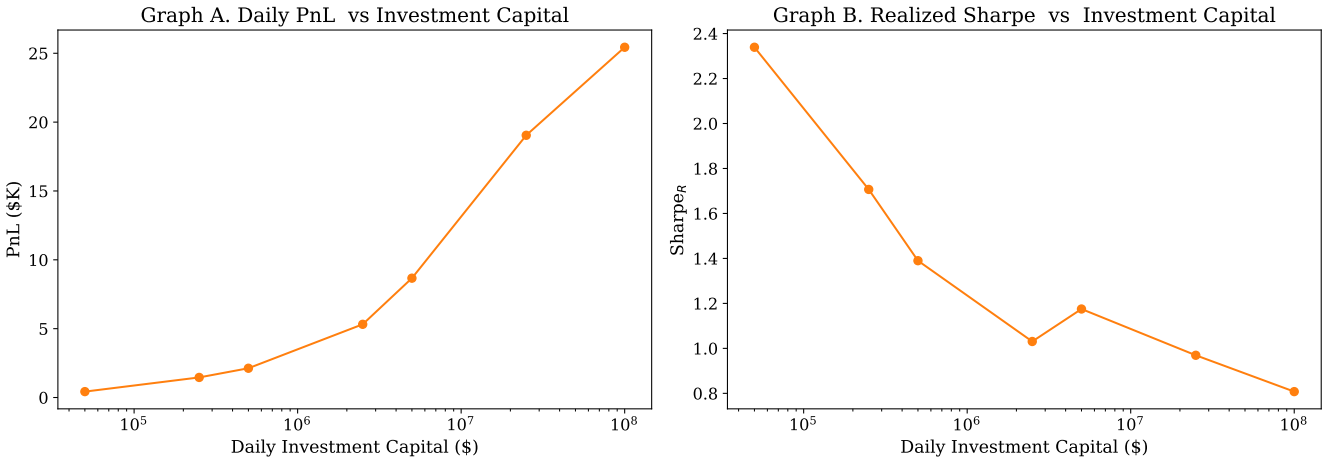
We employ an ensemble model with a penalized regressor to predict market excess returns. Using the baseline capacity parameters, we estimate the implementable portfolio value and apply the price-impact costs to account for trading costs. The experiment proceeds as follows.

On each trading day, stocks are ranked in descending order by the absolute value of their predicted returns. This ranking reflects trading priority because positions with larger expected payoffs (in absolute terms) are assumed to be more attractive to traders. Next, we consider six thresholds for the fund's daily investment capital, corresponding to different fund sizes ranging from \$50,000, to \$100,000,000. For each capital size level, trades are executed sequentially according to the ranking until the cumulative dollar trade size reaches the specified limit. Trades that exceed the capital limit are omitted. This procedure ensures that the strategy is feasible given the investment size of the fund.

Figure 10 presents the performance results. Panel A plots the average daily absolute PnL against investment capital, while Panel B reports the realized Sharpe ratio versus investment capital. Interestingly, the two graphs reveal opposing trends. As fund size increases, the average daily dollar return rises, and the realized Sharpe ratio declines.

Figure 10: Fund's Performance against Investment Capital

Performance of an investment fund as function of the investment size. We use predicted returns from the ensemble penalized regression model (Agg-PMBR) and impose different daily investment capital limits, which are proxies for fund size. We use the baseline capacity parameters and the price impact cost to estimate the implementable value and realized return for each stock. On each trading day, the strategy selects feasible trades in descending order of absolute predicted returns, subject to the constraint that the cumulative trade size does not exceed the specified investment capital limit. Panel A presents the results of daily realized dollar PnL under different investment levels, and Panel B reports the corresponding realized Sharpe ratio. The x-axis for both graphs is plotted on a logarithmic scale for ease of presentation.



At the early stage of a hedge fund, when investment capital is limited, traders tend to concentrate on positions with the highest expected returns, i.e., those associated with stronger predictive signals. Small investment level limits aggregate tradable volume and, hence, the dollar PnL. However, by focusing on a smaller set of high-confidence trades, the fund attains higher percentage returns and realized Sharpe ratio.

As the fund grows and its investment size expands, deploying larger amounts of capital requires moving further down the stock ranking to include securities with lower expected returns. While the dollar PnL continues to rise, it does so at a decreasing rate because additional capital is allocated to weaker signals. The x-axis in Figure 10 is on a logarithmic scale for ease of presentation; under a linear scale, the curve would appear concave. This pattern explains the declining trend in percentage returns as investment capital increases.

The pattern in Panel B is consistent with the well-documented diseconomies of scale, whereby growing fund size reduces investment efficiency (Grinblatt and Titman, 1989; Perold and Salomon Jr., 1991; Berk and Green, 2004; Chen et al., 2004; Yan, 2008; Aggarwal and Jorion, 2010). Chen et al. (2004) attribute this effect among mutual funds primarily to liquidity, and Yan (2008) document similar evidence for hedge funds.

Both studies find that the inverse relationship between performance and fund size is

strongest for portfolios concentrated in small and illiquid stocks. Consistent with this evidence, we observe that as investment capital expands, the average market value and average ADV of selected securities rise substantially.

Smaller funds may trade more nimbly (Aggarwal and Jorion, 2010), because their limited capacity allows greater flexibility to target stocks with higher expected returns. Many of these stocks are small and illiquid, yet they exhibit stronger predictability and higher confidence in forecasts.

In practice, large funds often impose trading restrictions for risk management purposes. Internal policies that set minimum thresholds for market value or ADV effectively exclude micro- and small-cap stocks from consideration. Consequently, the realized Sharpe ratio may decline further, because the tradable universe omits precisely those securities with the strongest predictability and highest percentage returns.

6 Conclusion

In this paper, we proposed a bottom-up framework to measure the implementable value of trading strategies. We showed that stock-level capacity limits constitute the primary trading contraints limiting the profitability of anomaly-based strategies.

Rather than relying on top-down portfolio construction under a unit-capital assumption, our bottom-up framework explicitly incorporates trading capacity limits at the individual-stock level. Once these constraints are considered, the Sharpe ratios of anomaly and return-prediction models decline substantially, and the tradable volume of most single-anomaly strategies is severely restricted, leaving only limited absolute profits to exploit.

Our results further reveal a divergence between predictability and profitability. Strong predictive performance is concentrated in illiquid small-cap stocks, where trading capacity is limited and the implementable scale is low. This finding carries an important implication for investors: in pursuing anomalies or models with higher predictive performance, it is equally important to evaluate the implementable value, because improvements in predictability often yield only negligible absolute gains in practice.

References

Abdi, F. and Ranaldo, A. (2017). A simple estimation of bid-ask spreads from daily close, high, and low prices. *Review of Financial Studies*, 30(12):4437–4480.

Aggarwal, R. K. and Jorion, P. (2010). The performance of emerging hedge funds and managers. *Journal of Financial Economics*, 96(2):238–256.

Almgren, R. F. (2003). Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10(1):1–18.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3):423–458.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3–18.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance*, 32(3):663–682.

Berk, J. B. and Green, R. C. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295.

Box, T., Davis, R., Evans, R., and Lynch, A. (2021). Intraday arbitrage between ETFs and their underlying portfolios. *Journal of Financial Economics*, 141(3):1078–1095.

Campbell, J. Y. and Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531.

Chen, A. Y. and Velikov, M. (2023). Zeroing in on the expected returns of anomalies. *Journal of Financial and Quantitative Analysis*, 58(3):968–1004.

Chen, A. Y. and Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review*, 27(2):207–264.

Chen, J., Hong, H., Huang, M., and Kubik, J. D. (2004). Does fund size erode mutual fund performance? the role of liquidity and organization. *The American Economic Review*, 94(5):1276–1302.

Chen, L., Pelger, M., and Zhu, J. (2024). Deep learning in asset pricing. *Management Science*, 70(2):714–750.

Chordia, T., Subrahmanyam, A., and Tong, Q. (2014). Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics*, 58(1):41–58.

Chung, K. H. and Zhang, H. (2014). A simple approximation of intraday spreads using daily data. *Journal of Financial Markets*, 17:94–120.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583.

Corwin, S. A. and Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *Journal of Finance*, 67(2):719–760.

DeMiguel, V., Martín-Utrera, A., Nogales, F. J., and Uppal, R. (2020). A transaction-cost perspective on the multitude of firm characteristics. *Review of Financial Studies*, 33(5):2180–2222.

Frazzini, A., Israel, R., and Moskowitz, T. J. (2012). Trading costs of asset pricing anomalies. Working paper.

Frazzini, A., Israel, R., and Moskowitz, T. J. (2018). Trading costs. Working paper.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis*. CRC Press, third edition.

Grinblatt, M. and Titman, S. (1989). Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Business*, 62(3):393–416.

Grinold, R. C. and Kahn, R. N. (1999). *Active portfolio management : a quantitative approach for providing superior returns and controlling risk*. McGraw-Hill, second edition.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273.

Guijarro-Ordonez, J., Pelger, M., and Zanotti, G. (2021). Deep learning statistical arbitrage. Working Paper.

Harvey, C. R., Liu, Y., and Zhu, H. (2015). ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68.

Haugen, R. A. and Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41(3):401–439.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer Nature.

Jahan-Parvar, M. R. and Zikes, F. (2023). When do low-frequency measures really measure effective spreads? evidence from equity and foreign exchange markets. *Review of Financial Studies*, 36(10):4190–4232.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance*, 45(3):881–898.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Jegadeesh, N. and Wu, Y. (2022). Closing auctions: Nasdaq versus NYSE. *Journal of Financial Economics*, 143(3):1120–1139.

Jesen, T. I., Kelly, B., and Pedersen, L. H. (2023). Is there a replication crisis in finance? *Journal of Finance*, 78(5):2465–2518.

Kishore, R., Brandt, M., Santa-Clara, P., and Venkatachalam, M. (2008). Earnings announcements are full of surprises. Working Paper.

Korajczyk, R. A. and Sadka, R. (2004). Are momentum profits robust to trading costs? *Journal of Finance*, 59(3):1039–1082.

Kyle, A. S. and Obizhaeva, A. A. (2016). Market microstructure invariance: Empirical hypotheses. *Econometrica*, 84(4):1345–1404.

Lesmond, D. A., Schill, M. J., and Zhou, C. (2004). The illusory nature of momentum profits. *Journal of Financial Economics*, 71(2):349–380.

Lo, A. W. and MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, 3(3):431–467.

Loughran, T. and Wellman, J. W. (2011). New evidence on the relation between the enterprise multiple and average stock returns. *Journal of Financial and Quantitative Analysis*, 46(6):1629–1650.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, 71(1):5–32.

Muravyev, D., Pearson, N. D., and Pollet, J. M. (2025). Anomalies and their short-sale costs. *The Journal of Finance*, 80(6):3639–3694.

Naranjo, A., Nimalendran, M., and Ryngaert, M. (1998). Stock returns, dividend yields, and taxes. *Journal of Finance*, 53(6):2029–2057.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1):1–28.

Novy-Marx, R. and Velikov, M. (2015). A taxonomy of anomalies and their trading costs. *Review of Financial Studies*, 29(1):104–147.

Park, C. (2005). Stock return predictability and the dispersion in earnings forecasts. *Journal of Business*, 78(6):2351–2376.

Perold, A. F. and Salomon Jr., R. S. (1991). The right amount of assets under management. *Financial Analysts Journal*, 47(3):31–39.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2009). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23(2):821–862.

Rapach, D. E. and Zhou, G. (2020). *Machine Learning for Asset Management*. John Wiley & Sons, Ltd.

Rasekhschaffe, K. C. and Jones, R. C. (2019). Machine learning for stock selection. *Financial Analysts Journal*, 75(3):pp. 70–88.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39(4):1127–1139.

Schwert, G. W. (2003). Anomalies and market efficiency. In *Financial Markets and Asset Pricing*, volume 1 of *Handbook of the Economics of Finance*, pages 939–974. Elsevier.

Tarun, C., Richard, R., and Avanidhar, S. (2008). Liquidity and market efficiency. *Journal of Financial Economics*, 87(2):249–268.

Timmermann, A. (2006). Forecast combinations. In *Handbook of Economic Forecasting*. Elsevier.

Yan, X. S. (2008). Liquidity, investment style, and the relation between fund size and fund performance. *Journal of Financial and Quantitative Analysis*, 43(3):741–767.

Zhao, A. B. and Cheng, T. (2022). Stock return prediction: Stacking a variety of models. *Journal of Empirical Finance*, 67:288–317.

A Models and Algorithms

A.1 Online Bayesian Regression

We follow the standard Bayesian estimation framework for linear regression, see Section 7 of Hoff (2009) and Section 14 of Gelman et al. (2013). Consider the standard linear model

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2),$$

where $\mathbf{x}_t \in \mathbb{R}^d$ are the input features (predictors), $y_t \in \mathbb{R}$ denotes the observed output (e.g. an asset return), $\boldsymbol{\beta} \in \mathbb{R}^d$ are the unknown regression coefficients, and ε_t is Gaussian noise. To obtain the Bayesian estimate of $\boldsymbol{\beta}$ at $t = 0$, we specify the prior belief

$$p(\boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)\,,$$

and given a new observation $(\mathbf{x}, y)$, we assume a Gaussian likelihood

$$p(y \mid \mathbf{x}, \boldsymbol{\beta}) = \mathcal{N}(y \mid \mathbf{x}^\top \boldsymbol{\beta}, \sigma^2)\,.$$

Applying Bayes' Rule, the posterior distribution of $\boldsymbol{\beta}$ is proportional to the product of the prior and the likelihood

$$\begin{aligned} p(\boldsymbol{\beta} \mid \mathbf{x}, y) \;&\propto\; p(\boldsymbol{\beta})\, p(y \mid \mathbf{x}, \boldsymbol{\beta}) \\ &\propto\; \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right) \exp\left(-\frac{1}{2\sigma^2}\,(y - \mathbf{x}^\top \boldsymbol{\beta})^2\right). \end{aligned}$$

Taking the logarithm of the posterior (up to a constant), we obtain

$$\begin{aligned} \log p(\boldsymbol{\beta} \mid \mathbf{x}, y) &= -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) \;-\; \frac{1}{2\sigma^2}\,(y - \mathbf{x}^\top \boldsymbol{\beta})^2 \;+\; \text{const} \\ &= -\frac{1}{2}\boldsymbol{\beta}^\top \left(\boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2}\,\mathbf{x}\,\mathbf{x}^\top\right) \boldsymbol{\beta} \;+\; \boldsymbol{\beta}^\top \left(\boldsymbol{\Sigma}_0^{-1}\,\boldsymbol{\mu}_0 + \frac{1}{\sigma^2}\,y\,\mathbf{x}\right) \;+\; \text{const}\,, \end{aligned}$$

which is the kernel of a multivariate normal distribution where

$$\text{Posterior covariance:} \quad \boldsymbol{\Sigma}_1 = \left(\boldsymbol{\Sigma}_0^{-1} \; + \; \frac{1}{\sigma^2} \, \mathbf{x} \, \mathbf{x}^\top \right)^{-1},$$

$$\text{Posterior mean:} \quad \boldsymbol{\mu}_1 = \boldsymbol{\Sigma}_1 \left(\boldsymbol{\Sigma}_0^{-1} \, \boldsymbol{\mu}_0 \; + \; \frac{1}{\sigma^2} \, y \, \mathbf{x} \right).$$

In an online setting, the posterior covariance and mean are updated sequentially as new data arrive. For notational simplicity, it is common to define the precision matrix as $\boldsymbol{P} = \boldsymbol{\Sigma}^{-1}$, which simplifies the update expressions and avoids repeated matrix inversions.

Time decay factor and sample weights. In the implementation of the online Bayesian regression algorithm, we incorporate a decay factor and sample weights to reflect better the dynamic and heterogeneous nature of streaming data. The decay factor, derived from a half-life parameter h, downweights the influence of older observations. Mathematically, one scales the prior precision by an exponential decay factor $\lambda = \exp(-\log 0.5/h)$ before multiplying the likelihood.

Analogous to weighted ordinary least squares, sample weights can also be incorporated into Bayesian regression. Specifically, before computing the sufficient statistics, both the predictor matrix $\mathbf{X}$ and the response vector $\mathbf{y}$ are pre-multiplied by the square root of the weights, yielding a weighted update that emphasizes more informative or reliable data points.

Bayesian estimation as loss minimization. In Bayesian decision theory, the optimal estimator is obtained by minimizing the expected loss

$$\widehat{\boldsymbol{\beta}} \; = \; \arg\min_{\boldsymbol{\beta}'} \, \mathbb{E}_{\boldsymbol{\beta} \sim p(\boldsymbol{\beta} | \mathcal{D})} \, [L(\boldsymbol{\beta}', \, \boldsymbol{\beta})]$$

under the posterior distribution, where $L(\boldsymbol{\beta}', \boldsymbol{\beta})$ denotes the loss incurred for estimating $\boldsymbol{\beta}'$ when the true value is $\boldsymbol{\beta}$. The form of the loss function L determines the corresponding Bayesian estimator. With squared loss $L(\boldsymbol{\beta}', \boldsymbol{\beta}) = \|\boldsymbol{\beta}' - \boldsymbol{\beta}\|^2$, the solution is the posterior mean $\widehat{\boldsymbol{\beta}} = \mathbb{E}[\boldsymbol{\beta} \mid \mathcal{D}]$.

In Bayesian linear regression with a Gaussian prior and Gaussian likelihood specifically, the posterior distribution is also Gaussian, and the posterior mean corresponds to the solution that minimizes a "regularized weighted least squares" loss function. As-

sume a dataset $\{(\mathbf{x}_t, y_t, w_t)\}_{t=1}^T$, where w_t are sample weights. The equivalent frequentist objective function is

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{t=1}^{T} w_t \left(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}\right)^2 \; + \; (\boldsymbol{\beta} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) \,,$$

where the first term corresponds to the weighted mean squared error as in weighted least squares regression. The second term reflects the influence of the Gaussian prior on $\boldsymbol{\beta}$, which can be interpreted as a quadratic regularization term. This loss function is the negative log-posterior, and minimizing it yields the Bayesian posterior mean.

Choice of initial prior. In Bayesian linear regression with a Gaussian prior, one typically starts with a non-informative or weakly informative prior. A common choice is a zero-mean Gaussian prior with an isotropic covariance matrix, i.e., $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathbf{I})$, where τ^2 is a large constant where in the paper we set $\tau^2 = 10^6$. This reflects the belief that coefficients are centered around zero but allows for high variance.

A.2 Predictive Models Combining the Anomalies

A.2.1 Multivariate Bayesian Regression

This model extends (11) to multiple predictors

$$r_{i,t} = \sum_{k=1}^{M} X_{i,t-1}^k \, \beta^k + \eta_t \;, \; i = 1, \ldots, N \,,$$

or, equivalently, in matrix form

$$\boldsymbol{r}_t = \boldsymbol{X}_{t-1} \, \boldsymbol{\beta} + \boldsymbol{\eta}_t \,, \tag{12}$$

where $\boldsymbol{r}_t \in \mathbb{R}^N$ denotes the vector of market excess returns in period $[t-1, t]$ for N assets, $\boldsymbol{\beta} \in \mathbb{R}^M$ is the vector of regression coefficients corresponding to M predictors, and $\boldsymbol{X}_t \in \mathbb{R}^{N \times M}$ is the design (predictor) matrix.

Analogous to the univariate case, let $\widehat{\boldsymbol{\beta}}_{t-1} \in \mathbb{R}^M$ and $\boldsymbol{\Lambda}_{t-1} \in \mathbb{R}^{M \times M}$ denote the prior mean and precision matrix (the inverse of the covariance matrix) of the coefficient vector.

At time t, the updates of the precision matrix and the new posterior mean are given by

$$\begin{aligned}\boldsymbol{\Lambda}_t &= \lambda\,\boldsymbol{\Lambda}_{t-1} \;+\; \boldsymbol{X}_{t-1}^{\top}\,\boldsymbol{W}_{t-1}\,\boldsymbol{X}_{t-1}\,,\\ \widehat{\boldsymbol{\beta}}_t &= \boldsymbol{\Lambda}_t^{-1}\left(\boldsymbol{\Lambda}_{t-1}\,\widehat{\boldsymbol{\beta}}_{t-1} \;+\; \boldsymbol{X}_{t-1}^{\top}\,\boldsymbol{W}_{t-1}\,\boldsymbol{r}_t\right).\end{aligned}$$

Here, λ is the decay factor and $\boldsymbol{W}$ is the diagonal weighting matrix, defined as in the univariate case. The estimate of the expected market excess return for the period $[t, t+1]$ is given by

$$\widehat{\boldsymbol{r}}_{t+1} = \boldsymbol{X}_t\,\widehat{\boldsymbol{\beta}}_t\,.$$

When new predictors are introduced, the dimensions of the posterior mean and covariance are expanded accordingly. The prior for each new coefficient is initialized to zero.

Although a simple linear model may suffer from overfitting in high-dimensional settings, it is often used as a benchmark for more complex models because it can be solved analytically.

A.2.2 Penalized Bayesian Regression

Similar to the extension from ordinary least squares to penalized regression, a regularization term can also be incorporated into Bayesian regression to mitigate overfitting. Standard ridge regression solves the regularized least squares problem

$$\widehat{\boldsymbol{\beta}} \;=\; \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} w_t\,\left(\,y_t - \mathbf{X}_t^{\top}\,\boldsymbol{\beta}\,\right)^2 \;+\; \lambda\,\|\boldsymbol{\beta}\|^2\,, \tag{13}$$

which is equivalent to maximum a posteriori estimation in Bayesian linear regression with a zero-mean Gaussian prior on $\boldsymbol{\beta}$. The strength of regularization is determined by the prior variance. Recall that the last term in the Bayesian loss function

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{t=1}^{T} w_t\left(y_t - \mathbf{x}_t^{\top}\boldsymbol{\beta}\right)^2 \;+\; \boldsymbol{\beta}^{\top}\,\boldsymbol{\Sigma}_0^{-1}\,\boldsymbol{\beta}$$

is a regularization term. Therefore, adjusting the value of τ^2 controls the strength of regularization: smaller values of τ^2 imply stronger shrinkage toward zero, equivalently to a larger ridge penalty $\lambda = \sigma^2/\tau^2$.

In the online updating process, this regularization is passed forward through the prior

precision matrix $\mathbf{P}_0 = \mathbf{\Sigma}_0^{-1}$. At each time step t, the update of the precision matrix is

$$\mathbf{P}_t = \mathbf{P}_{t-1} \; + \; \frac{1}{\sigma^2}\, \mathbf{x}_t\, \mathbf{x}_t^\top,$$

and the posterior mean becomes

$$\boldsymbol{\mu}_t = \mathbf{P}_t^{-1} \left(\mathbf{P}_{t-1}\, \boldsymbol{\mu}_{t-1} \; + \; \frac{1}{\sigma^2}\, y_t\, \mathbf{x}_t \right).$$

Thus, the effect of regularization, encoded in the initial prior, persists and influences the entire sequence of updates.

Accordingly, in Online Bayesian Ridge Regression, we initialize

$$\boldsymbol{\beta}_0 = \mathbf{0}\,, \quad \mathbf{\Sigma}_0 = \tau^2\, \mathbf{I} \quad \text{so that} \quad \boldsymbol{P}_0 = \frac{1}{\tau^2}\, \mathbf{I}\,,$$

where τ^2 controls the strength of the regularization imposed by the prior. At each time step t, given new observations $(\mathbf{X}_t, \mathbf{r}_t, \mathbf{w}_t)$, we perform updates identical in form to those in the standard Bayesian regression.

A.2.3 Boosted Regression Trees

Despite their simplicity and computational efficiency, linear models overlook the nonlinear effects of predictors and the interactions among them. Gu et al. (2020) point out that such interactions play an important role in explaining cross-sectional asset returns.

Decision trees provide a flexible alternative by recursively partitioning the data into subsets that are homogeneous in the target variable. Each split minimizes impurity, measured here by weighted mean squared error, and splitting continues until the maximum depth is reached. Predictions at the leaves are the average target values within those subsets.

We employ boosted regression trees, which combine many shallow trees into a strong ensemble. Each tree is trained to correct residual errors from prior trees by minimizing a loss function via gradient descent. Common implementations include XGBoost, LightGBM, and CatBoost: XGBoost emphasizes regularization and scalability; LightGBM uses histogram-based, leaf-wise growth for efficiency; and CatBoost handles categorical variables effectively while reducing overfitting through ordered boosting.

A.2.4 Model Ensemble

In addition to using individual models to generate stock return predictions, we explore approaches that ensemble the forecasts from multiple heterogeneous models. The idea of forecast combination arises in settings where multiple models produce predictions for the same target variable. Even if the best-performing model can be identified at each point in time, combining forecasts can yield diversification gains by exploiting complementary information across models to improve overall predictive performance and robustness (Timmermann, 2006).

Model ensembling is successfully applied across a wide range of empirical domains (Clemen, 1989), and more recently is used to predict stock returns. The literature evolved from simple averaging methods (Rapach et al., 2009) to more advanced ensembling techniques, including elastic net-based combinations (Rapach and Zhou, 2020) and general stacking frameworks (Zhao and Cheng, 2022).

In our analysis, we adopt two forecast aggregation methods: simple averaging and a (penalized) multivariate Bayesian regression.

Specifically, suppose there are total M different prediction models, let $\widehat{r}^{\,m}_{i,t+1}$ denote the one-step ahead return prediction for stock i, generated by model p at time t. The simple average forecast is computed as

$$\widehat{r}^{\,A}_{i,t+1} = \frac{1}{P} \sum_{m} \widehat{r}^{\,m}_{i,t+1} \,.$$

In the second approach, we specify the linear aggregation model

$$r_{i,t} = \sum_{p} \widehat{r}^{\,p}_{i,t} \; \beta^{p} \; + \; \eta_t \,,$$

where the weights β^m are updated sequentially with the penalized Bayesian regression procedure. Typically, forecasts from different models are highly correlated, so we adopt a ridge ensemble approach to mitigate multicollinearity effects. As mentioned earlier, we incorporate a penalty term λ and solve the regression coefficients in (13). Given the updated posterior mean estimated $\widehat{\beta}^{\,m}_{t}$ at time t, the aggregated forecast for the period $[t, t+1]$ is given by

$$\widehat{r}^{\,B}_{i,t+1} = \sum_{m} \widehat{r}^{\,m}_{i,t+1} \; \widehat{\beta}^{\,m}_{t} \,.$$

In this way, the combination weights adapt dynamically over time based on the evolving performance of individual models, offering a more flexible and data-driven forecast ensembling strategy.

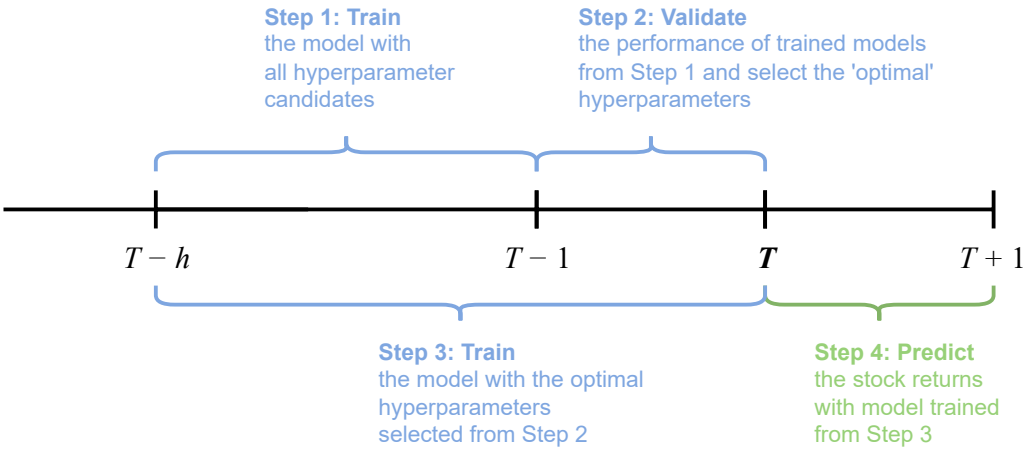
B Sample Splitting and Hyperparameter Tuning

B.1 Sample Splitting with Rolling Window

Except for the multivariate Bayesian regression, all predictive models used in the study require hyperparameter tuning. It is computationally expensive and almost unrealistic to update these models on a daily basis (as it is done with the univariate Bayesian regression), and hyperparameter tuning also requires enough data for training and validation to ensure the consistency and robustness of the model.

Figure 11 illustrates the sample splitting and tuning process with a rolling window of length h. Specifically, with the current time point defined as T, the period $[T - h, T - 1]$ is the training set to fit the model over a grid of hyperparameter candidates. Model performance is validated on the sample from $[T - 1, T]$, where the "optimal" hyperparameter set is selected. Finally, the model is retrained on the window $[T - h, T]$ with the chosen hyperparameters to forecast the excess return in the subsequent time period. In our analysis, we update the model annually and use a fixed rolling window of two years for all machine learning models.

Figure 11: Sample splitting and hyperparameter tuning steps



B.2 Hyperparameter Grid

Table 8: Hyperparameters For All Methods

The list of hyperparameters that we use to tune each machine learning model. For both penalized multivariate Bayesian regression (PMBR) and model ensembling using penalized regression (Agg_PMBR), we tune the prior for the covariance matrix to control the strength of regularization. For all three boosted regression trees (BRT), we tune the same set of hyperparamters including the maximum depth, the learning rates, the number of trees, and the column sampling ratio.

PMBR	**BRT**	**Agg_PMBR**
$\tau^2 \in \{10^{-6}, 10^{-3}, 1, 10^3, 10^6\}$	Depth $\in \{3, 5, 7\}$ Learning rate $\in \{0.1, 0.05, 0.01\}$ Trees $\in 75, 100, 150$ Column sampling $\in \{0.5, 0.8, 1.0\}$	$\tau^2 \in \{10^{-6}, 10^{-3}, 1, 10^3, 10^6\}$

C Transaction Costs

C.1 Half Bid-Ask Spread

Given the limited accessibility and availability of intraday quote data, many studies estimate transaction costs with low-frequency daily (Jahan-Parvar and Zikes, 2023). Chen and Velikov (2023) find that the average of low-frequency measures exhibits higher correlation with the high-frequency spread benchmark than any of the single proxy.

Here, we adopt the average of five popular low-frequency proxies: Roll (1984)'s estimate, Corwin and Schultz (2012)'s high-low estimate, Chung and Zhang (2014)'s CRSP spread estimate, Kyle and Obizhaeva (2016)'s volume-over-volatility measure, and Abdi and Ranaldo (2017)'s high-low-close estimate. To obtain a more stable measure, we further average the daily spread proxies over a rolling 30-day window and use this value as the estimate of the trading costs $C_{i,t}$ in (7) and (8).

C.2 Market Price Impact

We apply the market impact model proposed by Frazzini et al. (2018), in which price impact is modeled as a square root function of trade size, see also Grinold and Kahn (1999), Almgren (2003), and Jegadeesh and Wu (2022).

To estimate trading costs from price impact, we take the intended trade size from (6) and use

$$\log C_{i,t} = \beta \, \log(V_{i,t}/\mathrm{ADV}_{i,t}) \, + \, \alpha \, , \tag{14}$$

where $\beta = 0.5$ corresponds to the square-root functional form, and $\alpha = 1.70$ is the calibrated intercept parameter, see Frazzini et al. (2018).

The price impact cost estimate is not a perfect representation of trading costs. In practice, large orders are typically split into smaller trades which are executed gradually throughout the trading day to mitigate adverse price movements and the "walking-the-book" costs that would otherwise result from submitting the full order at once.

D Anomaly List

Table 9: List of Anomalies

List of 126 anomalies used in the paper, including their corresponding acronyms, economic categories, and update frequencies. All 105 monthly predictors are taken from Chen and Zimmermann (2022), using the August 2024 data release available at `https://www.openassetpricing.com/`. The 21 daily predictors are constructed from the CRSP daily database.

Acronym	Description	Economic Category	Frequency
AM	Total assets to market	valuation	monthly
Accruals	Accruals	accruals	monthly
AnnouncementReturn	Earnings announcement return	earnings event	monthly
AssetGrowth	Asset growth	investment	monthly
BM	Book to market, original (Stattman 1980)	valuation	monthly
BMdec	Book to market using December ME	valuation	monthly
BPEBM	Leverage component of BM	leverage	monthly
Beta	CAPM beta	risk	monthly
BetaLiquidityPS	Pastor-Stambaugh liquidity beta	liquidity	monthly
BetaTailRisk	Tail risk beta	risk	monthly
BookLeverage	Book leverage (annual)	leverage	monthly
CBOperProf	Cash-based operating profitability	profitability	monthly
CF	Cash flow to market	valuation	monthly
Cash	Cash to assets	asset composition	monthly
CashProd	Cash Productivity	profitability alt	monthly
ChAssetTurnover	Change in Asset Turnover	sales growth	monthly
ChEQ	Growth in book equity	investment	monthly
ChInv	Inventory Growth	investment alt	monthly
ChNNCOA	Change in Net Noncurrent Op Assets	investment alt	monthly
ChNWC	Change in Net Working Capital	investment alt	monthly
ChTax	Change in Taxes	other	monthly
CompEquIss	Composite equity issuance	external financing	monthly
CompositeDebtIssuance	Composite debt issuance	external financing	monthly
ConvDebt	Convertible debt indicator	external financing	monthly
CoskewACX	Coskewness using daily returns	risk	monthly
Coskewness	Coskewness	risk	monthly

Continued on next page

Acronym	Description	Economic Category	Frequency
DelCOA	Change in current operating assets	investment alt	monthly
DelCOL	Change in current operating liabilities	external financing	monthly
DelEqu	Change in equity to assets	investment	monthly
DelFINL	Change in financial liabilities	external financing	monthly
DelLTI	Change in long-term investment	investment	monthly
DelNetFin	Change in net financial assets	investment alt	monthly
DolVol	Past trading volume	volume	monthly
EBM	Enterprise component of BM	valuation	monthly
EP	Earnings-to-Price Ratio	valuation	monthly
EntMult	Enterprise Multiple	valuation	monthly
ExclExp	Excluded Expenses	composite accounting	monthly
FEPS	Analyst earnings per share	profitability	monthly
ForecastDispersion	EPS Forecast Dispersion	volatility	monthly
GP	gross profits / total assets	profitability	monthly
GrLTNOA	Growth in long term operating assets	investment	monthly
GrSaleToGrInv	Sales growth over inventory growth	sales growth	monthly
GrSaleToGrOverhead	Sales growth over overhead growth	sales growth	monthly
Herf	Industry concentration (sales)	other	monthly
HerfAsset	Industry concentration (assets)	other	monthly
HerfBE	Industry concentration (equity)	other	monthly
High52	52 week high	momentum	monthly
IdioVol3F	Idiosyncratic risk (3 factor)	volatility	monthly
IdioVolAHT	Idiosyncratic risk (AHT)	volatility	monthly
Illiquidity	Amihud's illiquidity	liquidity	monthly
IndMom	Industry Momentum	momentum	monthly
IndRetBig	Industry return of big firms	lead lag	monthly
IntMom	Intermediate Momentum	momentum	monthly
IntanBM	Intangible return using BM	long term reversal	monthly
IntanCFP	Intangible return using CFtoP	long term reversal	monthly
IntanEP	Intangible return using EP	long term reversal	monthly
InvGrowth	Inventory Growth	profitability	monthly
InvestPPEInv	change in ppe and inv/assets	investment	monthly
Investment	Investment to revenue	investment	monthly
LRreversal	Long-run reversal	long term reversal	monthly
Leverage	Market leverage	leverage	monthly
MaxRet	Maximum return over month	volatility	monthly
Mom12m	Momentum (12 month)	momentum	monthly
Mom12mOffSeason	Momentum without the seasonal part	other	monthly
Mom6m	Momentum (6 month)	momentum	monthly
MomOffSeason	Off season long-term reversal	other	monthly
MomOffSeason06YrPlus	Off season reversal years 6 to 10	other	monthly
MomSeason	Return seasonality years 2 to 5	other	monthly
MomSeason06YrPlus	Return seasonality years 6 to 10	other	monthly

Continued on next page

Acronym	Description	Economic Category	Frequency
MomSeasonShort	Return seasonality last year	other	monthly
NOA	Net Operating Assets	asset composition	monthly
NetDebtFinance	Net debt financing	external financing	monthly
NetEquityFinance	Net equity financing	external financing	monthly
OPLeverage	Operating leverage	other	monthly
OperProfRD	Operating profitability R&D adjusted	profitability	monthly
PctAcc	Percent Operating Accruals	accruals	monthly
PctTotAcc	Percent Total Accruals	accruals	monthly
Price	Price	other	monthly
PriceDelayRsq	Price delay r square	lead lag	monthly
RDS	Real dirty surplus	composite accounting	monthly
REV6	Earnings forecast revisions	earnings forecast	monthly
RealizedVol	Realized (Total) Volatility	volatility	monthly
ResidualMomentum	Momentum based on FF3 residuals	momentum	monthly
ReturnSkew	Return skewness	risk	monthly
ReturnSkew3F	Idiosyncratic skewness (3F model)	risk	monthly
SP	Sales-to-price	valuation	monthly
STreversal	Short term reversal	short-term reversal	monthly
ShareIss1Y	Share issuance (1 year)	external financing	monthly
ShareVol	Share Volume	volume	monthly
Size	Size	size	monthly
Tax	Taxable income to income	other	monthly
TotalAccruals	Total accruals	investment alt	monthly
TrendFactor	Trend Factor	momentum	monthly
VolSD	Volume Variance	liquidity	monthly
XFIN	Net external financing	external financing	monthly
betaVIX	Systematic volatility	volatility	monthly
cfp	Operating Cash flows to price	valuation	monthly
dNoa	change in net operating assets	investment	monthly
grcapx	Change in capex (two years)	investment growth	monthly
hire	Employment growth	investment alt	monthly
roaq	Return on assets (qtrly)	profitability	monthly
std_turn	Share turnover volatility	liquidity	monthly
zerotrade12M	Days with zero trades	liquidity	monthly
zerotrade1M	Days with zero trades	liquidity	monthly
zerotrade6M	Days with zero trades	liquidity	monthly
5d_volatility	5-day realized volatility	volatility	daily
MV	Market value	size	daily
PRC	Price	other	daily
VOL	share volume	volume	daily
beta_prev	CAPM beta	risk	daily
bid_ask_spread	Bid-ask spread	liquidity	daily
Ret1d	1 day past return	short term reversal	daily

Continued on next page

Acronym	Description	Economic Category	Frequency
illiq	Amihud's illiquidity	liquidity	daily
liq	trading volume	volume	daily
liq_std	trading volume volatility	liquidity	daily
log(MV)	log market value	size	daily
log(PRC)	log price	other	daily
log(liq)	log trading volume	volume	daily
log(liq_std)	log trading volume volatility	liquidity	daily
log(turnover)	log share turnover	liquidity	daily
log(turnover_std)	log share turnover volatility	liquidity	daily
Ret1M	1 month past return	short term reversal	daily
Ret1W	1 week past return	short term reversal	daily
Ret3M	3 month past return	long term reversal	daily
turnover	share turnover	liquidity	daily
turnover_std	share turnover volatility	liquidity	daily

E Single Anomaly Performance Results

Table 10 and 11 present the performance evaluation for each single anomaly during the IS and OOS periods respectively. Both tables report predictability metrics: the weighted R^2 (3), the hypothetical Sharpe_H and the t-statistic t_H (5); and the profitability metrics under capacity constraints: dollar trade size S_t (6), realized dollar PnL_t (8), realized percentage return Ψ_t (9), realized Sharpe ratio Sharpe_R and the corresponding t-statistic of realized return t_R (10). The baseline capacity parameters are $\{\delta = 1\%, \phi = 5\%, \text{Cap} = \$1\text{M}\}$, and trading cost are set to zero, i.e., $C_{i,t} = 0$.

Table 10: Single Anomalies' In-Sample Performance

Performance evaluation for each single anomaly during the IS period. The table includes the R^2 (%) (3), hypothetical Sharpe ratio Sharpe_H and the t-statistic t_H (5), dollar trade size S_t (\$K) (6), realized dollar PnL_t (\$) (8), realized percentage return Ψ_t (%) (9), realized Sharpe ratio Sharpe_R and the corresponding t-statistic of realized return t_R (10). The baseline capacity parameters are $\{\delta = 1\%, \phi = 5\%, \text{Cap} = \$1\text{M}\}$, and trading cost are set to zero, i.e., $C_{i,t} = 0$.

Anomaly acronym	R^2	Sharpe_H	t_H	Daily PnL	Daily S	Sharpe_R	t_R	Ψ
AM	−0.0066	0.03	0.16	−23	22	0.12	0.63	0.08
Accruals	0.0321	0.88	4.74	321	83	0.70	3.77	0.27
AnnouncementReturn	0.1017	2.45	9.45	1,199	285	1.40	5.42	0.38
AssetGrowth	0.0270	1.39	8.21	3,506	242	0.88	5.23	0.69
BM	0.1286	0.97	3.63	55	34	0.24	0.91	0.20
BMdec	0.0001	0.35	1.83	17	4	0.01	0.07	0.01
BPEBM	−0.0005	0.03	0.18	14	0	−0.21	−1.28	−0.15
Beta	−0.0346	−0.06	−0.33	−10	2	−0.23	−1.31	−0.16
BetaLiquidityPS	0.0003	0.16	0.91	−19	114	0.07	0.38	0.03
BetaTailRisk	−0.0091	0.02	0.14	−970	406	−0.23	−1.55	−0.15
BookLeverage	−0.0017	0.10	0.52	3	2	0.17	0.88	0.16

Continued on next page

Acronym	R^2	Sharpe$_H$	t_H	Daily PnL	Daily S	Sharpe$_R$	t_R	Ψ
CBOperProf	0.1029	1.64	11.70	13,322	2,863	0.63	4.53	0.32
CF	0.0221	0.80	3.74	136	50	0.46	2.15	0.24
Cash	−0.0195	0.00	0.03	−4,893	818	−0.19	−1.17	−0.16
CashProd	−0.0016	0.19	1.18	−30	11	0.32	2.00	0.39
ChAssetTurnover	−0.0024	−0.17	−0.71	121	23	0.13	0.56	0.21
ChEQ	0.0035	0.61	3.97	1,370	109	0.61	4.00	0.78
ChInv	0.0389	1.06	5.51	1,090	221	0.92	4.77	0.39
ChNNCOA	0.0042	0.93	3.92	−133	603	0.41	1.76	0.14
ChNWC	−0.0007	0.26	1.09	−426	163	0.44	1.86	0.24
ChTax	0.0019	0.79	4.23	−163	339	0.18	0.95	0.15
CompEquIss	−0.0187	−0.07	−0.43	−951	157	−0.03	−0.20	−0.04
CompositeDebtIssuance	0.0041	0.72	4.25	332	425	0.33	1.98	0.12
ConvDebt	0.0066	1.00	5.21	2,336	2,666	0.33	1.74	0.11
CoskewACX	0.0221	0.25	1.54	3,225	479	0.37	2.30	0.21
Coskewness	0.0088	0.11	0.60	217	246	−0.09	−0.50	−0.04
DelCOA	0.0405	0.79	4.95	469	523	0.39	2.42	0.24
DelCOL	0.0044	0.35	2.17	−540	290	0.07	0.45	0.05
DelEqu	0.0355	0.60	3.70	820	518	0.39	2.39	0.40
DelFINL	0.0316	1.21	7.57	1,625	444	0.41	2.59	0.16
DelLTI	−0.0004	0.14	0.87	32	43	0.04	0.25	0.03
DelNetFin	0.0131	0.99	6.16	551	351	0.08	0.51	0.03
DolVol	−0.0241	−0.20	−1.08	−5	112	−0.09	−0.51	−0.05
EBM	−0.0005	−0.03	−0.20	13	1	−0.28	−1.72	−0.21
EP	0.0481	0.19	0.71	6	10	−0.24	−0.90	−0.22
EntMult	−0.0000	0.32	2.20	−61	43	0.11	0.74	0.20
ExclExp	−0.0023	−0.22	−0.71	−285	108	−0.31	−1.04	−0.12
FEPS	−0.0027	0.54	2.35	2,268	649	0.52	2.28	0.35
ForecastDispersion	−0.0028	0.80	3.91	−84	139	0.47	2.33	0.24
GP	0.0209	0.67	4.58	3,793	1,203	0.23	1.59	0.11
GrLTNOA	0.0071	0.51	2.74	110	47	0.41	2.18	0.16
GrSaleToGrInv	−0.0032	1.04	3.90	26	7	0.83	3.10	0.33
GrSaleToGrOverhead	−0.0261	0.49	1.84	150	78	0.51	1.91	0.21
Herf	−0.0000	0.22	1.37	−105	218	0.12	0.76	0.05
HerfAsset	−0.0008	0.12	0.72	−198	168	0.00	0.01	0.00
HerfBE	−0.0006	−0.12	−0.75	−16	13	−0.11	−0.68	−0.06
High52	0.1111	0.79	4.84	5,134	1,307	0.61	3.77	0.41
IdioVol3F	0.1197	1.02	6.18	2,909	1,386	0.49	3.00	0.25
IdioVolAHT	0.3180	2.35	10.75	6,845	1,879	2.19	10.05	0.45
Illiquidity	0.1180	0.66	3.81	375	561	0.41	2.34	0.21
IndMom	−0.0296	0.19	1.08	−14	188	0.21	1.17	0.14
IndRetBig	0.5021	1.44	7.73	3,631	688	0.44	2.39	0.30
IntMom	−0.0302	0.02	0.19	613	345	0.21	1.83	0.21
IntanBM	−0.0137	−0.11	−0.62	−2,299	302	−0.27	−1.62	−0.17

Continued on next page

Acronym	R^2	Sharpe$_H$	t_H	Daily PnL	Daily S	Sharpe$_R$	t_R	Ψ
IntanCFP	−0.0011	0.26	1.56	−1,550	315	0.00	0.00	0.00
IntanEP	−0.0036	0.06	0.33	−1,700	276	−0.07	−0.43	−0.06
InvGrowth	0.0041	0.89	5.91	113	21	0.77	5.10	0.53
InvestPPEInv	0.0135	1.27	7.48	1,169	234	0.90	5.33	0.43
Investment	−0.0012	0.54	2.57	128	59	0.35	1.69	0.11
LRreversal	−0.0641	0.13	0.86	−29	9	−0.08	−0.57	−0.07
Leverage	−0.0104	−0.20	−1.10	−6	2	−0.29	−1.54	−0.17
MaxRet	0.1022	0.81	5.32	7,459	1,871	0.29	1.91	0.22
Mom12m	0.0640	0.68	3.37	188	170	0.58	2.91	0.59
Mom12mOffSeason	0.0008	0.56	3.37	4,049	1,182	0.44	2.66	0.47
Mom6m	0.0677	0.55	2.74	39	172	0.43	2.16	0.37
MomOffSeason	0.0560	0.58	3.53	1,385	1,053	0.30	1.84	0.26
MomOffSeason06YrPlus	0.0161	0.60	3.65	−941	392	0.16	0.99	0.09
MomSeason	0.0153	0.48	2.89	4,503	631	0.44	2.69	0.27
MomSeason06YrPlus	0.0276	0.59	3.57	3,074	449	0.56	3.42	0.27
MomSeasonShort	−0.0130	0.22	1.33	−2,651	440	0.08	0.48	0.07
NOA	0.0267	1.47	9.07	1,913	283	1.02	6.27	0.42
NetDebtFinance	0.0318	1.35	7.14	596	358	0.52	2.77	0.17
NetEquityFinance	0.0727	1.49	7.86	3,583	730	1.03	5.46	0.44
OPLeverage	−0.0007	0.14	0.91	−465	182	0.01	0.08	0.01
OperProfRD	0.0396	0.93	6.62	7,126	1,343	0.40	2.85	0.23
PctAcc	−0.0009	0.15	0.66	−155	50	−0.12	−0.54	−0.09
PctTotAcc	−0.0006	0.32	1.39	153	34	0.22	0.97	0.18
Price	−0.0076	0.41	2.46	57	9	0.77	4.61	0.70
PriceDelayRsq	0.0048	0.31	1.91	−3,446	902	−0.39	−2.38	−0.22
RDS	−0.0021	0.66	3.41	740	308	0.25	1.31	0.18
REV6	0.0250	1.38	5.35	213	58	0.86	3.32	0.31
RealizedVol	0.1227	1.06	6.43	2,298	1,296	0.41	2.52	0.27
ResidualMomentum	0.0962	1.12	9.61	6,226	1,691	0.89	7.61	0.44
ReturnSkew	−0.0036	0.12	0.82	−200	339	0.06	0.42	0.02
ReturnSkew3F	−0.0017	0.14	0.99	1,516	567	0.51	3.57	0.17
SP	−0.0082	0.31	1.06	−45	105	0.02	0.07	0.01
STreversal	−0.0810	0.16	1.18	−28	9	−0.23	−1.69	−0.19
ShareIss1Y	0.0316	1.89	10.85	4,062	395	1.39	7.98	0.57
ShareVol	0.0135	0.39	2.11	280	136	−0.05	−0.25	−0.05
Size	−0.0071	0.31	1.99	47	9	0.84	5.33	0.68
Tax	−0.0010	0.01	0.03	−31	7	−0.00	−0.02	−0.00
TotalAccruals	0.0051	0.62	3.88	796	103	0.68	4.26	0.40
TrendFactor	−0.0340	0.10	0.89	−3,070	972	0.09	0.79	0.07
VolSD	−0.0204	0.50	2.67	−184	343	0.18	0.97	0.13
XFIN	0.0764	1.63	8.63	836	736	0.80	4.24	0.30
betaVIX	0.0039	0.53	2.00	1,923	497	0.31	1.16	0.16
cfp	0.0123	1.18	5.78	146	52	0.87	4.24	0.30

Continued on next page

Acronym	R^2	Sharpe$_H$	t_H	Daily PnL	Daily S	Sharpe$_R$	t_R	Ψ
dNoa	0.0243	1.44	8.84	1,564	167	0.81	5.01	0.53
grcapx	0.0017	1.00	4.79	156	21	0.74	3.53	0.52
hire	0.0099	0.41	2.74	705	1,163	0.29	1.92	0.19
roaq	0.0643	1.34	7.21	4,880	1,037	0.69	3.70	0.34
std_turn	−0.0251	0.64	3.42	16	6	0.18	0.98	0.25
zerotrade12M	0.0045	0.30	1.96	−1,781	609	−0.30	−1.99	−0.21
zerotrade1M	−0.0065	0.05	0.34	−1,564	346	−0.34	−2.26	−0.24
zerotrade6M	0.0020	0.22	1.42	−1,655	562	−0.32	−2.07	−0.22
5d_volatility	−0.0090	0.50	3.02	−2	957	−0.02	−0.10	−0.00
MV	0.0002	0.69	4.52	−0	6	−0.01	−0.08	−0.00
PRC	0.0009	0.46	2.87	−0	12	−0.01	−0.03	−0.00
VOL	0.0066	3.25	12.17	533	1,936	0.85	3.18	0.02
beta_prev	−0.0022	2.03	4.76	7	22	1.36	8.12	0.07
bid_ask_spread	0.0443	2.27	9.87	52	192	0.49	2.13	0.02
Ret1d	0.0582	1.95	14.44	189	166	0.13	0.98	0.02
illiq	−0.0006	0.79	4.53	−93	214	0.09	0.53	0.00
liq	0.0011	1.84	9.89	9	892	0.26	1.39	0.01
liq_std	−0.0007	0.26	1.38	−29	325	−0.22	−1.21	−0.01
log(MV)	0.0020	0.70	4.55	−2	11	0.46	3.03	0.03
log(PRC)	0.0040	0.84	5.26	−2	11	−0.10	−0.64	−0.01
log(liq)	0.0546	4.97	27.19	356	3,691	0.75	4.11	0.02
log(liq_std)	0.0010	0.82	4.39	24	1,387	0.19	1.01	0.00
log(turnover)	0.0176	2.49	13.42	183	649	0.95	5.10	0.03
log(turnover_std)	0.0031	0.82	4.41	−57	664	0.08	0.42	0.00
Ret1M	0.0184	1.44	10.65	7	106	0.08	0.59	0.01
Ret1W	0.0896	2.24	16.58	38	140	0.33	2.46	0.03
Ret3M	−0.0174	0.51	3.79	−9	68	0.07	0.49	0.01
turnover	0.0304	4.73	25.46	374	621	1.00	5.36	0.06
turnover_std	−0.0005	0.54	2.90	−42	110	−0.20	−1.06	−0.01

Table 11: Single Anomalies' Out-of-Sample Performance

Performance evaluation for each single anomaly during the IS period. The table includes the R^2 (%) (3), hypothetical Sharpe ratio Sharpe$_H$ and the t-statistic t_H (5), dollar trade size S_t (\$K) (6), realized dollar PnL$_t$ (\$) (8), realized percentage return Ψ_t (%) (9), realized Sharpe ratio Sharpe$_R$ and the corresponding t-statistic of realized return t_R (10). The baseline capacity parameters are $\{\delta = 1\%, \phi = 5\%, \text{Cap} = \$1\text{M}\}$, and trading cost are set to zero, i.e., $C_{i,t} = 0$.

Anomaly acronym	R^2	Sharpe$_H$	t_H	Daily PnL	Daily S	Sharpe$_R$	t_R	Ψ
AM	−0.0321	0.06	0.36	−7,512	2,019	0.07	0.42	0.06
Accruals	−0.0000	0.23	1.30	681	1,462	0.37	2.15	0.16
AnnouncementReturn	0.0215	1.09	6.13	8,835	6,589	0.46	2.60	0.18
AssetGrowth	−0.0180	0.76	3.48	2,723	1,609	0.29	1.32	0.17
BM	−0.0004	0.64	4.41	−2,499	4,889	−0.00	−0.01	−0.00

Continued on next page

Acronym	R^2	Sharpe$_H$	t_H	Daily PnL	Daily S	Sharpe$_R$	t_R	Ψ
BMdec	−0.0005	−0.01	−0.05	−77	46	0.10	0.59	0.06
BPEBM	−0.0012	−0.05	−0.23	−12	8	0.15	0.70	0.17
Beta	0.0721	0.70	5.21	18,770	6,822	0.51	3.81	0.33
BetaLiquidityPS	−0.0097	−0.18	−0.92	−3,805	1,442	−0.07	−0.34	−0.03
BetaTailRisk	−0.0176	0.11	0.42	9,275	8,022	0.15	0.55	0.06
BookLeverage	−0.0011	−0.08	−0.48	−3	86	−0.18	−1.07	−0.15
CBOperProf	0.0172	0.83	2.61	16,209	7,027	0.89	2.81	0.26
CF	0.0062	0.48	2.80	7,476	4,977	0.40	2.35	0.21
Cash	−0.0403	0.05	0.21	−954	4,139	0.16	0.62	0.09
CashProd	0.0007	0.22	0.99	713	35	0.33	1.50	0.52
ChAssetTurnover	−0.0021	−0.13	−0.59	−236	26	−0.65	−3.04	−1.04
ChEQ	−0.0012	−0.37	−1.52	−25	98	−0.40	−1.65	−0.32
ChInv	0.0010	0.45	2.32	3,877	2,922	0.33	1.69	0.15
ChNNCOA	−0.0119	0.19	0.89	2,767	710	0.33	1.55	0.10
ChNWC	−0.0019	0.29	1.34	971	1,201	0.29	1.36	0.09
ChTax	−0.0004	−0.11	−0.45	−34	3	−0.16	−0.67	−0.15
CompEquIss	−0.0203	−0.28	−1.27	−2,066	992	−0.06	−0.25	−0.05
CompositeDebtIssuance	−0.0041	0.23	1.02	402	2,828	−0.11	−0.48	−0.04
ConvDebt	−0.0007	0.27	0.92	3,257	3,630	0.11	0.38	0.06
CoskewACX	−0.0598	0.20	0.97	−3,540	11,756	−0.07	−0.33	−0.04
Coskewness	−0.0372	0.07	0.38	−1,496	3,992	0.06	0.31	0.03
DelCOA	−0.0152	−0.09	−0.45	−2,547	1,516	−0.02	−0.12	−0.02
DelCOL	−0.0136	0.02	0.10	−555	1,319	−0.12	−0.55	−0.07
DelEqu	−0.0298	0.36	1.72	2,944	1,500	0.16	0.77	0.11
DelFINL	0.0056	0.47	2.23	5,835	4,644	0.37	1.76	0.13
DelLTI	−0.0017	−0.55	−2.65	110	351	−0.35	−1.68	−0.17
DelNetFin	−0.0052	0.12	0.56	−1,663	2,911	−0.03	−0.15	−0.01
DolVol	−0.0310	−0.21	−1.11	−16,077	4,517	−0.28	−1.48	−0.18
EBM	−0.0011	−0.11	−0.54	−38	10	0.09	0.41	0.08
EP	0.0023	0.28	2.07	−490	518	−0.10	−0.72	−0.06
EntMult	−0.0252	0.07	0.28	2,645	796	−0.09	−0.35	−0.10
ExclExp	−0.0017	0.52	2.59	5,663	4,009	0.31	1.54	0.12
FEPS	−0.0556	0.42	1.98	774	87	0.44	2.04	0.29
ForecastDispersion	−0.0126	0.47	2.30	1,265	722	0.30	1.45	0.15
GP	0.0174	0.87	3.24	20,308	8,570	0.83	3.09	0.28
GrLTNOA	−0.0056	−0.01	−0.03	−489	797	0.13	0.74	0.05
GrSaleToGrInv	−0.0016	−0.15	−0.93	−43	7	−0.51	−3.04	−0.51
GrSaleToGrOverhead	−0.0399	0.45	2.68	794	167	0.37	2.19	0.25
Herf	−0.0265	0.22	1.05	697	2,715	0.04	0.21	0.01
HerfAsset	−0.0207	0.28	1.34	235	3,174	0.08	0.37	0.02
HerfBE	−0.0017	0.03	0.14	−242	58	−0.11	−0.55	−0.12
High52	−0.0894	0.35	1.65	18,944	14,064	0.25	1.18	0.14
IdioVol3F	−1.7041	0.73	3.57	27,920	22,131	0.30	1.49	0.16

Continued on next page

Acronym	R^2	Sharpe$_H$	t_H	Daily PnL	Daily S	Sharpe$_R$	t_R	Ψ
IdioVolAHT	0.0024	0.43	2.21	18,696	17,686	0.14	0.73	0.09
Illiquidity	−0.1036	0.12	0.63	−14,931	4,647	−0.18	−0.95	−0.12
IndMom	−0.1070	−0.34	−1.84	−24,860	3,714	−0.51	−2.73	−0.45
IndRetBig	0.0072	0.53	2.53	9,324	3,579	0.37	1.75	0.24
IntMom	0.0008	−0.02	−0.07	5,818	5,531	−0.06	−0.21	−0.03
IntanBM	−0.0148	−0.06	−0.26	385	2,211	0.10	0.44	0.05
IntanCFP	−0.0092	0.15	0.70	−3,226	2,269	−0.19	−0.87	−0.13
IntanEP	−0.0047	0.19	0.87	−310	2,961	−0.00	−0.01	−0.00
InvGrowth	−0.0130	0.10	0.37	512	532	−0.07	−0.28	−0.05
InvestPPEInv	−0.0357	0.01	0.05	−69	998	−0.13	−0.56	−0.11
Investment	−0.0017	0.50	2.66	−314	813	0.19	1.02	0.10
LRreversal	0.0033	0.31	1.98	1,985	1,944	0.32	2.07	0.25
Leverage	−0.0037	−0.12	−0.81	−2,309	657	−0.23	−1.50	−0.20
MaxRet	0.0118	0.73	3.15	3,079	9,881	0.12	0.51	0.05
Mom12m	−0.0428	0.38	2.26	7,005	4,098	0.29	1.72	0.33
Mom12mOffSeason	−0.0905	−0.11	−0.52	−10,066	3,881	0.00	0.00	0.00
Mom6m	−0.0749	0.25	1.50	−648	2,680	0.14	0.83	0.15
MomOffSeason	−0.0401	0.07	0.33	1,480	3,632	0.10	0.47	0.07
MomOffSeason06YrPlus	0.0025	0.47	2.17	3,270	3,505	0.21	0.96	0.09
MomSeason	−0.0511	−0.06	−0.26	−2,404	2,326	0.03	0.13	0.01
MomSeason06YrPlus	−0.0038	0.36	1.68	6,997	3,548	0.61	2.86	0.23
MomSeasonShort	−0.0988	−0.06	−0.28	−1,222	3,466	0.04	0.19	0.02
NOA	−0.0073	0.01	0.05	2,715	884	0.19	0.87	0.09
NetDebtFinance	0.0055	0.63	3.09	6,175	4,623	0.43	2.09	0.14
NetEquityFinance	−0.2161	0.97	4.72	31,578	10,830	0.93	4.55	0.35
OPLeverage	0.0038	0.66	2.64	12,403	6,403	0.64	2.57	0.19
OperProfRD	0.0359	1.07	3.37	21,775	8,161	0.86	2.70	0.26
PctAcc	−0.0003	−0.04	−0.16	27	52	0.28	1.11	0.14
PctTotAcc	−0.0005	−0.21	−0.83	−165	28	0.04	0.16	0.03
Price	0.0349	0.72	5.20	7,886	8,902	0.39	2.81	0.19
PriceDelayRsq	−0.0263	0.06	0.27	−7,135	5,751	−0.11	−0.53	−0.05
RDS	−0.0010	0.09	0.40	−530	605	0.03	0.12	0.03
REV6	0.0007	0.53	2.98	502	218	0.49	2.79	0.35
RealizedVol	−2.0996	0.68	3.34	20,296	21,698	0.25	1.24	0.17
ResidualMomentum	−0.0253	0.08	0.31	−1,404	4,784	0.07	0.28	0.03
ReturnSkew	−0.0009	−0.15	−0.53	−831	2,234	−0.10	−0.33	−0.02
ReturnSkew3F	−0.0022	0.16	0.56	−2,926	2,927	−0.28	−0.96	−0.06
SP	−0.0153	0.09	0.50	−3,821	1,683	−0.10	−0.57	−0.07
STreversal	−0.0570	0.12	0.72	−4,951	2,261	0.16	0.96	0.13
ShareIss1Y	0.0002	0.66	3.03	154	93	0.44	2.03	0.20
ShareVol	−0.0222	−0.08	−0.44	−5,868	2,335	−0.20	−1.13	−0.15
Size	0.0027	0.45	3.14	1,937	5,738	0.35	2.45	0.16
Tax	−0.0058	−0.31	−1.51	−1,005	64	−0.27	−1.34	−0.24

Continued on next page

Acronym	R^2	Sharpe$_H$	t_H	Daily PnL	Daily S	Sharpe$_R$	t_R	Ψ
TotalAccruals	−0.0103	0.12	0.58	−395	361	0.15	0.70	0.09
TrendFactor	−0.1162	−0.43	−1.35	−13,839	10,568	−0.50	−1.56	−0.23
VolSD	−0.0047	0.33	1.75	−1,088	3,447	0.03	0.19	0.03
XFIN	−0.1509	1.04	5.10	47,219	10,959	0.98	4.79	0.41
betaVIX	−0.3314	0.23	1.10	5,797	4,891	0.36	1.77	0.21
cfp	0.0097	0.61	3.17	7,779	3,065	0.27	1.42	0.20
dNoa	−0.0070	0.38	1.79	2,317	1,088	0.13	0.63	0.08
grcapx	−0.0034	0.16	0.82	−97	41	0.16	0.82	0.18
hire	0.0040	0.60	2.24	7,784	4,702	0.23	0.85	0.11
roaq	−0.0178	0.38	1.64	4,487	2,955	0.53	2.31	0.23
std_turn	0.0104	0.99	5.32	2,224	119	0.89	4.76	1.85
zerotrade12M	−0.0119	0.13	0.59	592	2,775	0.09	0.42	0.04
zerotrade1M	−0.0124	−0.12	−0.54	1,973	3,334	0.10	0.45	0.04
zerotrade6M	−0.0117	0.14	0.63	1,290	3,050	0.08	0.36	0.03
5d_volatility	−0.0021	1.02	5.30	1,237	36,324	0.38	1.99	0.02
MV	0.0000	0.56	2.90	1,199	6,501	0.35	1.84	0.01
PRC	−0.0000	−0.39	−2.05	−24	100	−0.69	−3.57	−0.03
VOL	−0.0013	0.24	1.25	330	20,605	0.10	0.54	0.01
beta_prev	−0.0110	0.09	0.45	1,487	32,137	0.17	0.89	0.01
bid_ask_spread	−0.0001	0.26	1.35	4	259	0.33	1.70	0.01
Ret1d	0.0273	1.56	8.11	11,672	37,905	0.18	0.92	0.01
illiq	−0.0000	0.06	0.33	−538	2,538	−0.46	−2.37	−0.01
liq	−0.0004	0.04	0.20	−1,113	7,725	−0.04	−0.21	−0.00
liq_std	−0.0006	−0.00	−0.01	−1,043	10,281	−0.10	−0.54	−0.01
log(MV)	−0.0020	0.22	1.15	−3,717	31,117	−0.22	−1.12	−0.01
log(PRC)	−0.0013	0.41	2.15	−2,862	26,963	−0.07	−0.38	−0.00
log(liq)	0.0040	0.90	4.67	−828	54,605	0.24	1.25	0.01
log(liq_std)	−0.0034	0.18	0.95	−5,579	45,902	−0.29	−1.52	−0.01
log(turnover)	0.0030	0.71	3.70	−1,817	27,920	0.16	0.82	0.01
log(turnover_std)	−0.0002	0.54	2.81	−997	23,859	0.13	0.68	0.01
Ret1M	−0.0026	1.01	5.24	−3,632	29,926	0.31	1.60	0.02
Ret1W	0.0248	1.77	9.21	21,626	44,317	0.76	3.97	0.04
Ret3M	−0.0064	0.39	2.05	−8,476	22,087	−0.12	−0.62	−0.01
turnover	−0.0045	0.43	2.23	−213	4,454	−0.07	−0.35	−0.01
turnover_std	−0.0001	0.86	4.48	1,037	2,385	0.55	2.88	0.04

F Performance Results on Large-Cap Stocks

We perform the analysis on large-cap stocks as a robustness check. Each year, we compute the average market value of each stock and select those in the top quintile (top 20%) by market value. For each single anomaly, we refit the univariate model in (11) using this restricted stock universe and then compute the predictability and profitability metrics within the bottom-up framework.

Table 12 reports the average performance of anomalies under the baseline capacity parameters, and Figure 12 visualizes the performance for each single-anomaly based strategies. The overall pattern is similar to that observed for the full sample in Table 1. In the IS period, incorporating capacity constraints reduces the Sharpe ratio by an average of 19% , while in the OOS period the decline is 8% ($p = 5.1 \times 10^{-4}$).

The magnitude of the reduction from Sharpe_H to Sharpe_R is smaller when the analysis is restricted to large-cap stocks, reflecting their generally higher trading capacity. Nevertheless, the decline remains statistically significant. Although capacity limits are less restrictive among large-cap stocks, substantial heterogeneity in ADV persists even within this group. Capacity constraints continue to represent a non-negligible limitation on strategy scalability.

Moreover, despite the relatively mild capacity constraints among large-cap stocks, trading signals are not necessarily strong enough to justify large trading volumes. Figure 13 plots average daily dollar returns against dollar trading volume: most single-anomaly strategies still operate at modest scale and generate limited dollar profits, despite exhibiting statistically significant percentage returns. Consequently, the bottom-up framework remains a more realistic and informative approach for evaluating the economic contribution of anomaly-based strategies.

Table 12: Average Anomaly's Performance among Large-Cap Stocks

The average performance of single-anomaly strategies applied to large-cap stocks in the top quintile (top 20%) by market value. Panel A reports predictability metrics: the weighted R^2 in % (3), hypothetical Sharpe ratio Sharpe_H and the t-statistic t_H (5). Panel B presents the profitability metrics: dollar trade size S_t in \$K (6), realized dollar PnL_t in \$ (8), realized percentage return Ψ_t in % (9), realized Sharpe ratio Sharpe_R and the corresponding t-statistic of realized return t_R (10). The baseline capacity parameters are $\{\delta = 1\%, \phi = 5\%, \text{Cap} = \$1\text{M}\}$, and trading costs are set to zero, i.e., $C_{i,t} = 0$.

	IS	OOS
Panel A. Predictability		
Average R^2	0.03	−0.11
Average Sharpe_H	0.61	0.24
Number of anomalies with $t_H > 1.96$	75	37
Ratio of anomalies with $t_H > 1.96$	0.59	0.29
Panel B. Profitability under capacity constraints		
Average daily dollar volume traded S_t	210	3,500
Average daily PnL_t	880	4,200
Average daily realized return Ψ_t	0.14	0.08
Average Sharpe_R	0.37	0.18
Number of anomalies with $t_R > 1.96$	61	27
Ratio of anomalies with $t_R > 1.96$	0.48	0.21

Figure 12: Single Anomaly's Sharpe_H Against Sharpe_R among Large-Cap Stocks

Hypothetical Sharpe ratio, Sharpe_H, of each anomaly against its corresponding realized Sharpe ratio, Sharpe_R, for strategies applied to large-cap stocks in the top quintile (top 20%) by market value. Panel A displays IS results, while Panel B displays OOS results. Each scatter point represents the performance of one anomaly. Both panels include fitted zero-intercept regression lines. Benchmark lines are drawn at $\text{Sharpe}_C = 0.5$ and $\text{Sharpe}_R = 0.5$ as rough indicators for meaningful performance in terms of Sharpe ratios.

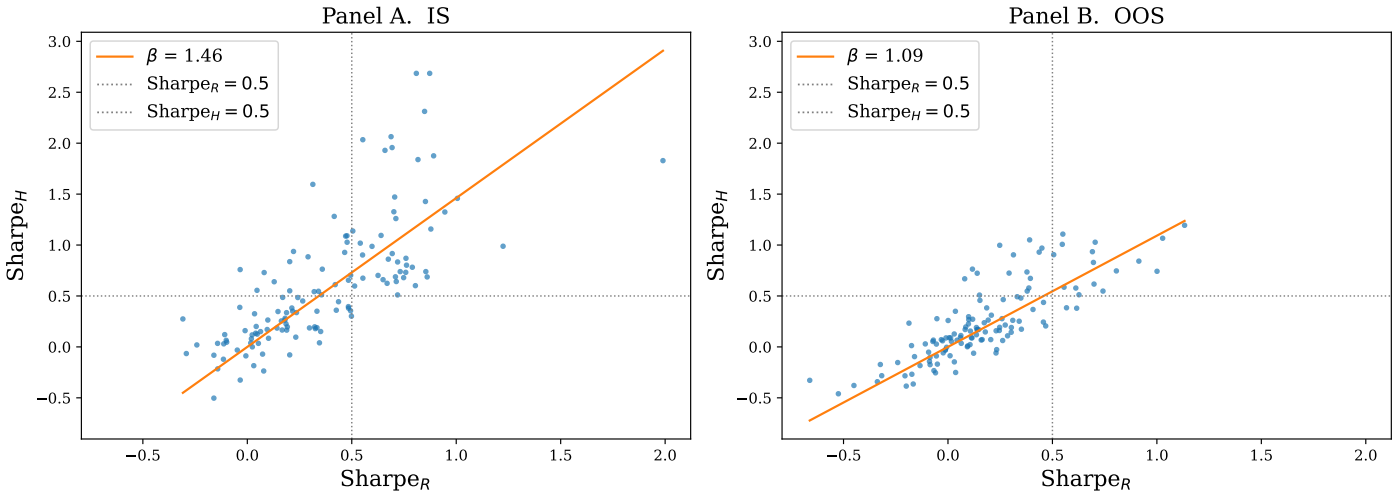


Figure 13: Single Anomaly's PnL against Volume Traded among Large-Cap Stocks

Scatter plots of the average daily dollar PnL, see (8), against the corresponding average daily dollar volume traded S, see (6), for strategies applied to large-cap stocks in the top quintile (top 20%) by market value. Each point represents one anomaly-based strategy. Panel A displays IS results, and Panel B displays OOS results. The color intensity of the scatter points reflects the anomaly's realized Sharpe_R, where darker points indicate higher realized Sharpe values. This visualizes the link among dollar volume traded, realized dollar PnL, and risk-adjusted percentage return.