Categorical Thinking and Experience Effect: Evidence from Price-Based Return Predictability

by*

Chun Liu Huaixin Wang Yuehan Wang Jianfeng Yu

January 2026

Abstract

Price-based categorical thinking and experience effects jointly imply that investor demand for a stock would be higher after similarly priced stocks earned high returns in the past. We show that when a focal stock's price-comparable peers experienced high (low) returns in the prior month, the focal stock earns abnormally high (low) returns subsequently. A long-short strategy that sorts stocks on the average return of price-comparable peers delivers 1.13% per month. Exploiting stock splits as a quasi-natural experiment that reshapes price categories and disrupts experience associations, we show that the price-based return predictability declines sharply post-split. These findings suggest the unique role of categorical thinking and the experience effect in shaping investor tastes and attention.

JEL Classification: G12, G41

Keywords: Categorical Thinking, Experience Effect, Style Investing, Predictability

---

* First Draft: January 2026. Preliminary and incomplete, please do not circulate. All errors are our own. Author affiliation/contact information: Chun Liu is with the School of Economics and Management, Tsinghua University: liuch@sem.tsinghua.edu.cn; Huaixin Wang is with the Faculty of Business Administration, University of Macau: huaixinwang@um.edu.mo; Yuehan Wang is with the School of Finance, Central University of Finance and Economics: wangyh@cufe.edu.cn; Jianfeng Yu is with the PBC School of Finance, Tsinghua University: yujf@pbcsf.tsinghua.edu.cn.

1 Introduction

Earlier studies such as Barberis and Shleifer (2003) argue that investors tend to think categorically based on firm characteristics. Different investors might also exhibit distinct preferences for stock characteristics. For example, retail investors tend to prefer low-priced stocks (e.g., Kumar and Lee (2006)), whereas institutional investors prefer high-priced stocks (e.g., Gompers and Metrick (2001)). More recently, Balasubramaniam et al. (2023) document a strong investor clientele effect associated with share prices. In a related study, Green and Hwang (2009) find that stocks that undergo splits experience an increase in comovement with low-priced stocks and a decrease in comovement with high-priced stocks, suggesting investors categorize stocks based on price. In addition, the psychology literature argues that personal experience substantially shapes individual decisions(Weber et al., 1993; Hertwig et al., 2004) and implies that investors with positive (negative) prior investment experiences are more (less) likely to purchase new stocks with similar characteristics or styles. Indeed, Huang (2019) finds that after experiencing a positive return from a stock holding, investors' propensity to buy other stocks with similar characteristics such as size, book-to-market, and industry also increases.

This paper investigates the asset pricing implications of jointly considering price-based categorical thinking and the experience effect. In particular, we posit that after experiencing a high return on their stock investment, investors could pay more attention to similarly priced stocks due to price-based categorical thinking and experience effect. Thus, high returns on the currently held stock attract investor attention, which subsequently spills over to similarly priced stocks. As a result, this behavior would lead to higher effective demand for stocks with comparable prices and, in turn, higher future returns for these stocks. The above arguments imply that if the peer group of similarly priced stocks related to a focal stock has realized high returns in the past, total demand for this focal stock tends to rise, and thus the subsequent return should be higher as well.

We define price-comparable peer stocks based on nominal share prices and examine the predictive ability of price-based peers' average returns for the focal stock's return. We find that after similarly priced stocks have experienced high (low) returns in the past month, the focal stock subsequently tends to earn abnormally high (low) returns. A long-short strategy that sorts stocks on lagged one-month average return of their price-comparable peers earns a monthly CAPM alpha of 1.27% (t=6.58) and a Fama and French (2015) five-factor alpha of 1.31% (t=5.91). This result remains robust after controlling for a wide range of factors and

firm-level characteristics using both double sorted portfolio analysis and Fama and MacBeth (1973) regressions.

Unlike traditional categorical thinking based on size or book-to-market, which may proxy for fundamentals or risk exposures, price-based categorical thinking appears to be difficult to reconcile with a purely rational framework since prices can be altered arbitrarily by splits and reverse splits. These practices render firm-level nominal prices largely devoid of economic meaning. Thus, it is hard to rationalize the price-based return predictability within a purely risk-based story. To provide further evidence on our proposed mechanism, we investigate stock price-based return predictability around stock splits. This setting provides a quasi-natural experiment for testing whether price-based categorization and the experience effect jointly exert significant return predictability. Stock splits induce large mechanical changes in nominal prices without corresponding changes in firm fundamentals. Therefore, they offer a relatively clean setting for testing the asset pricing implications of category-based investing with few confounding effects. To proceed, we examine the shifts in split stocks' return predictability with price-indexed portfolios before and after stock splits.

We first perform an excise similar to (Green and Hwang, 2009) by replacing their contemporaneous comovement regressions with predictive regressions. If investors categorize stocks by price, then after a split we would expect that low-priced stocks have stronger predictive power for the focal than high-priced stocks. Consistent with this hypothesis, we find that stocks undergoing splits exhibit a decline in the predictive power of high-priced stocks for the focal stock and an increase in the predictive power of low-priced stocks. Actually, following splits, low-priced stocks exert stronger predictive power than high-priced stocks.

We then turn to cross-sectional price-based return predictability explored in this paper. In particular, our mechanism relies on the *interaction* of two ingredients: (1) categorical thinking by price (Green and Hwang, 2009), and (2) anchoring on past investment experience (Huang, 2019). Both are necessary, and neither channel alone can generate our result. More specifically, since individuals tend to anchor to prior experiences, their portfolio choices adjust sluggishly over time (Campbell et al., 2014; Kuhnen et al., 2017; Bazley et al., 2021). A stock split mechanically reduces the nominal share price, creating a wedge between the stock's new price-based peer group and the peer group that investors have mentally associated with the stock. Consistent with this implication, we document an immediate decline in price-based return predictability following split events. To further isolate the mechanism, we implement a difference-in-differences design that contrasts 2-for-1 splits (a 50% price reduction) with

moderate-factor splits that imply smaller price changes (e.g., 3-for-2 and 5-for-4). This design helps mitigate concerns that split decisions reflect unobserved fundamentals (Shue and Townsend, 2021). In line with our story, price-based return predictability declines more following larger nominal price cuts (i.e., 2-for-1 splits). These results highlight the importance of experience-based anchoring in generating return predictability among price-comparable stocks.

To validate the role of price-based categorical thinking and complement our analysis, we conduct two placebo tests. First, we show that the return predictability does not exhibit systematic shifts after splits when peers are defined along alternative characteristics, such as industry, size, book-to-market, profitability, or investment, rather than nominal price. As such, the traditional style investing (Barberis and Shleifer, 2003) alone cannot explain our findings. Second, we examine variations in price-based return predictability around alternative corporate events (Kwon and Tang, 2025). In contrast to stock splits, these events do not induce systematic changes in predictability.

Our evidence on stock splits supports the hypothesis that the cross-sectional return predictability we document is primarily driven by price-based categorization interacting with experience effect. We also assess whether the traditional style investing (Barberis and Shleifer, 2003) contributes to this pattern using standard empirical tests. Prior studies (e.g., Teo and Woo (2004) and Froot and Teo (2008)) report positive-feedback trading at the style level; Lewellen (2002) and Wahal and Yavuz (2013) find momentum effects in characteristic-sorted portfolios, plausibly attributable to such style-level feedback trading. To rule out the possibility that our price-based return predictability merely reflects the well-known style-investing effect, we estimate Fama and MacBeth (1973) regressions that explicitly control for style returns (as in Wahal and Yavuz (2013)) and for industry returns across multiple horizons. After these controls, the average past return of price-comparable peers remains a strong predictor of focal stocks' subsequent returns.

To ensure the price-based return predictability is not an artifact of other known anomalies, we conduct a broad set of robustness checks. First, we perform double-sorting portfolio analysis and Fama-MacBeth regressions to control for a large set of firm-level characteristics, including size, book-to-market, past returns, profitability, investment, idiosyncratic volatility, illiquidity, and skewness. We find that our price-based return predictability remains significant after all of these controls. Second, we conduct a placebo test by redefining peers using randomly matched stocks. We show that the predictive power of the resulting variable indeed disappears. Lastly, we extend our analysis to 38 international

markets, and obtain broadly consistent results.

From an asset pricing perspective, the price-based return predictability may reflect initial underreaction, subsequent overreaction, or a combination of both, analogous to momentum (Jegadeesh and Titman, 1993). On the one hand, strong performance among price-based peers may induce unjustified demand for the focal stock, thereby generating continued price overreaction during the post-formation period. Price-based categorical thinking interacting with experience effect implies such a channel. On the other hand, albeit less likely, stocks at similar price levels may share common fundamentals. If recent high returns among price-based peers convey favorable information about the focal stock's fundamentals, then the focal stock should subsequently earn abnormal returns because it was initially underpriced at portfolio formation, potentially due to limited investor attention. While our tests based on stock splits provide support for the continued overreaction hypothesis, we further inspect the mechanism by examining long-horizon return predictability and subsequent order imbalance.

Following the approach of Jegadeesh and Titman (1993), we construct price-peer return-based trading strategies with long holding periods. It turns out that the price-based return predictability completely reverses 30 months after portfolio formations. The strategy return even becomes negative at longer horizons. This long-term reversal pattern aligns with the continued overreaction story and iis difficult to reconcile with an initial underreaction explanation.

We proceed to investigate subsequent retail order imbalance for portfolios sorted on the recent performance of price-comparable peers. Stocks whose peers earned high recent returns experience significantly larger increases in retail order imbalance than stocks whose peers performed poorly, indicating stronger retail demand following favorable peer performance. This evidence complements Huang (2019), who show that individuals tend to purchase stocks sharing similar size and book-to-market after personally experiencing high returns.[1] the trading response is markedly weaker for large orders, where institutional investors are more likely to dominate. Taken together, these patterns suggest that experience-driven excess demand, rather than rational arbitrage, more plausibly underlies our price-based return predictability.

Parallel to trading-based implications, price-based categorical thinking interacting with

[1]Whereas Huang (2019) study repeat category purchases by the same investors, the order-imbalance measure aggregates retail trading across all investors and does not require prior ownership of similarly priced peers.

experience effects also generates testable predictions for investor attention. If investors make decisions around price categories, then recent performance among price-based peers should capture attention and spill over to stocks the investor currently holds. To test this prediction, we use abnormal Google search volume (Da et al., 2011) to proxy for retail investors' attention. In a series of panel regressions, we find a strong positive association between the magnitude of price peers' lagged returns and retail attention. Decomposing peer returns into positive and negative components, we further show that both favorable and unfavorable experiences significantly influence attention allocation.

Lastly, we investigate heterogeneity in the price-based return predictability. The effect is statistically significant across large and small firms, across high- and low-idiosyncratic volatility stocks, high- and low-institutional ownership stocks, and in both early and recent periods. Whereas price-based return predictability is stronger among smaller firms, high-volatility stocks, low-institutional ownership stocks, and in the earlier sample, the pattern aligns with the notion that short-sale constraints impede arbitrage in retail-dominated stocks (see, e.g., Nagel (2005)). Overall, these results indicate that price-based return predictability intensifies where retail participation is greater and arbitrage is more difficult.

Our paper relates closely to Green and Hwang (2009), who find that investors tend to categorize stocks based on price, and they show that stocks that undergo splits experience an increase in comovement with low-priced stocks and a decrease in comovement with high-priced stocks. Whereas Green and Hwang (2009) study the contemporaneous comovement, we examine the *predictive* power of similarly priced stocks for the focal stock. Our paper is also highly related Huang (2019), who finds that after investors personally experience positive returns from their stock holdings, they become more likely to purchase other stocks in the same industry and stocks with similar size and book-to-market. While Huang (2019) emphasize investor trading behavior, we focus on the asset-pricing implications of the experience effect interacting with price-based categorical thinking, namely, the cross-sectional predictive power of the average returns of price-comparable stocks for the focal stock. We show that price-based style investing generates economically large alphas for portfolio strategies formed on the average returns of similarly priced stocks, even after controlling for a broad set of firm-level characteristics.

In addition, our paper contributes to the broader literature on style investing and categorical thinking. Early work on style investing (e.g., Barberis and Shleifer (2003) and Barberis et al. (2005)) argues that, to simplify portfolio choice, many investors first sort

stocks into categories (such as small-value stocks or high-tech firms) and then allocate capital at the category level rather than at the individual-security level. Such capital allocation can generate excess comovement among firms within the same category. Peng and Xiong (2006) provide a theoretical foundation for categorical thinking based on attention constraints, and a large subsequent literature documents that category investing affects price formation, particularly through excess comovement. Our story differs from traditional style investing in an important way. The price-based return predictability we document does not require investors to trade at the category level; investors may still invest at the individual-stock level. Instead, we only requires that, following high recent returns, investors become more likely to purchase new stocks with similar prices. In fact, our price-based return predictability remains statistically and economically significant after controlling for the standard style effects documented in prior studies. More important, unlike firm size and book-to-market, price is less economically meaningful and is less likely to proxy for systematic risk. This feature further distinguishes our results from existing evidence.

Furthermore, our study relates to the literatures on investor attention and positive-feedback trading. A large body of work has found a significant role for attention and positive feedback trading in asset pricing anomalies.[2] As Barber and Odean (2008) emphasize, trading requires attention, and investors confront a substantial search problem when selecting stocks to purchase. We posit that, following a high return for a given stock, investors tend to pay more attention to stocks with comparable prices because of price-based categorical thinking or a price-based clientele effect. This attention reallocation increases demand and, consequently, predicts higher subsequent returns for stocks whose similarly priced peers have recently performed well. The positive feedback channel also implies that investors tend to increase their positions after a positive investment experience (Pearson et al., 2021). In fact, investors tend to trade more intensively after a positive investment experience, even if the positive experience comes from winning IPO lotteries purely by chance (Ben-David et al., 2018; Anagol et al., 2021; Gao et al., 2021). Relatedly, Chen et al. (2023) show that attention can spill over to firms with neighboring stock codes after a positive experience. Distinct from these studies, we argue that after observing positive returns on a focal stock, investors become especially likely to trade stocks at similar price levels, despite the absence of an obvious economic linkage among firms that share similar nominal prices.

Lastly, our paper connects to the emerging literature on the role of memory in financial

[2]This literature is too voluminous to summarize here, but readers can see Daniel and Hirshleifer (2015), Barber et al. (2019), and Gabaix (2019) for reviews.

decision-making (e.g., Bordalo et al. (2025a,b); Wachter and Kahana (2024)). After an investor experiences high returns on a stock investment, that experience may heighten the cognitive accessibility of price-comparable stocks through associative memory, thereby shifting preferences toward such stocks. In addition, a positive experience may increase subsequent demand for similarly priced stocks since the past gain potentially serves as a cue that triggers memory retrieval and mental simulation of high payoffs, amplifying investors' appetite for other similarly priced stocks. Consequently, prior exogenous investment gains can systematically distort subsequent portfolio choices through memory-based belief formation. Our setting provides a unique test of these memory-based implications by using nominal price as a salient associative cue that links past experiences to future portfolio choices.

The rest of the paper is organized as follows. We describe data sources and variable definitions in Section 2. Section 3 presents our main results using portfolio analysis and Fama-MacBeth regressions. Section 4 examines the mechanism underlying the predictability results. Section 5 further explores several additional robustness tests. Section 6 concludes.

2 Data Description and Summary Statistics

Our sample includes non-financial ordinary common stocks listed on the NYSE, AMEX, and NASDAQ (CRSP exchange codes 1 to 3; share codes 10 and 11) from 1963 to 2024. Stock return and price data are from the Center for Research in Security Prices (CRSP), and accounting data are from Compustat. Institutional ownership is obtained from the Thomson Financial 13F filings. Trade-level transaction data come from the Institute for the Study of Securities Markets (ISSM) for 1983 to 1992 and the Trade and Quote (TAQ) database for 1993 to 2000. Data on non-U.S. firms are sourced from Compustat Global. To mitigate the influence of microcaps, our main analysis relies on value-weighted returns. To alleviate the effect of a bid-ask bounce, we also follow Campbell et al. (2008) to delete stocks with a price less than \$1 at the portfolio formation date.

2.1 Definition of Key Predictor

The primary variable of interest is the average return of a stock's price-based peers. At the end of each month t, we standardize nominal share prices cross-sectionally to have zero mean

and unit variance. For each focal stock i, we compute its distance to every other stock j as

$$\text{Distance}_{i,j,t} = |\text{Price}_{i,t} - \text{Price}_{j,t}|. \tag{1}$$

We define price-based peers as the 100 stocks with the smallest $\text{Distance}_{i,j,t}$. We then calculate the value-weighted average monthly return of these price-based peers, denoted by PSIM.

To mitigate the traditional short-term reversal effect (Jegadeesh, 1990; Lehmann, 1990; Lo and MacKinlay, 1990) and to address the concern that peer returns may merely proxy for investor attention to extreme returns of the focal stock, we follow Chen et al. (2023) and orthogonalize the peer-stock average return with respect to the focal stock's own return over the same window used to compute PSIM. This procedure yields a residual variable, denoted RPSIM, which serves as the key measure in our main analysis.[3] Importantly, RPSIM isolates the variation in similarly priced peers' returns that is orthogonal to the focal stock's performance, rather than conflating peer returns with short-term reversal effects. For robustness, we also report results using the unadjusted PSIM measure.

2.2 Control Variables

To assess whether our price-based return predictability is subsumed by other established anomalies, we consider a large number of firm-level characteristics, including (1) Size, measured as the natural logarithm of market capitalization; (2) book-to-market ratio (BM); (3) Profitability (Profit), measured by return-on-assets; (4) Investment (Invest), measured by growth in total assets; (5) Medium-term momentum (Mom), defined as the cumulative return from month $t-12$ to $t-2$; (6) Long-term reversal (Ltrev), defined as the cumulative return from month $t-60$ to $t-13$; (7) Idiosyncratic volatility (IVol) as in Ang et al. (2006); (8) Maximum daily return (MAX) within the month, as in Bali et al. (2011); (9) Skewness (Skew), computed as the skewness of monthly returns over the prior five years; and (10) Coskewness (Coskew), following Harvey and Siddique (2000) and estimated using five years of monthly returns.

In addition, we include style-related controls based on size, book-to-market, and industry. We follow Wahal and Yavuz (2013) to construct size-BM style returns. Specifically, at the

[3]Specifically, in each month we run a cross-sectional regression with PSIM as the dependent variable and the focal stock's corresponding monthly return as the regressor. We define RPSIM as the regression residual.

end of each June, we form 5×5 portfolios using quintile breakpoints for size and book-to-market. NYSE size breakpoints are obtained from Kenneth French's website. We define the style return for a given size-BM cell as the value-weighted average return of stocks in that cell. For industry styles, we compute the value-weighted average return of stocks within each Fama-French 48-industry classification.

2.3 International Sample

Following Gao et al. (2018), our international sample includes common stocks listed on 43 stock exchanges across 38 areas/countries. We convert all returns and accounting variables to U.S. dollars and apply sample screens at the area/country level. In a given month (or week), we include an area/country only if at least 50 stocks have non-missing return observations. Following Hou et al. (2011), we further require each stock to have at least 12 monthly observations over the sample period. To alleviate the influence of small cap firms, we exclude micro-cap firms that have a price or market capital below 5% in each month (week) in an area/country. Following Ince and Porter (2006), we also implement two filters to mitigate the impact of extreme returns. First, we treat as missing any return exceeding 300% that reverses within one month. Specifically, if R_t or R_{t-1} exceeds 300% and $(1+R_t)(1+R_{t-1})-1 < 50\%$, then we set both R_t and R_{t-1} to missing. Second, we set returns to missing if they fall outside the 0.1st and 99.9th percentiles of the return distribution. The international sample spans 1997 to 2024.

2.4 Summary Statistics

At the beginning of each month, we sort stocks into decile portfolios based on each stock's 100 most similarly priced stocks' residual average return in the past month. We then hold the portfolio for a month and then rebalance it at the beginning of each month. Table 1 reports various characteristics for portfolios formed on RPSIM. Specifically, in each month we compute, for each decile, the cross-sectional mean of a set of firm characteristics across the stocks in that decile, and we then report time-series averages of these monthly decile means. The reported variables include the sorting variable RPSIM, the next-month value of the sorting variable (RPSIM_{t+1}), the natural logarithm of market capitalization (Size), share price (Prc), market beta (Beta), the book-to-market ratio (BM), profitability (Profit), investment (Invest), idiosyncratic volatility (IVol), the focal stock's past one-month return

(R1m), the past 12-month return excluding the most recent month (Mom), and the Amihud (2002) illiquidity measure (Illiq).

Overall, high-RPSIM stocks exhibit slightly lower betas, slightly larger market capitalization, higher price level, higher book-to-market ratios, modestly higher profitability, lower investment rates, lower idiosyncratic volatility, and lower illiquidity.[4] The last row of Table 1 reports time-series averages of the cross-sectional correlations between RPSIM and other firm characteristics. We notice that RPSIM is positively correlated with firm size and share price. These correlations are unlikely to drive our predictability results, however, because prior studies document negative return predictability for size and price, whereas our price-based return predictability implies a positive relation with RPSIM. Our findings remain robust after controlling for size and price premia. We also report the subsequent one-month peer-return measure for each decile (RPSIM_{t+1}). Notably, RPSIM_{t+1} is higher for the high-RPSIM portfolio, a pattern that contrasts sharply with the standard short-term reversal effect.

3 The Predictive Power of Price-Comparable Stock Returns

In this section, we first perform our main portfolio analysis based on our key variable RPSIM. Then we use double-sorting and Fama-MacBeth regression analysis to control for other compounding effects. We also study the long-horizon return behavior after portfolio formation to tease out potential underlying mechanisms.

3.1 Unconditional Sorts

Table 2 presents our main portfolio evidence using unconditional sorts on RPSIM. Each month, stocks are grouped into decile portfolios based on RPSIM. Portfolios are held for one-month and we calculate value-weighted and equal-weighted returns. Panel A shows that stocks whose price-comparable peers earned high returns in the previous month subsequently outperform those whose price-comparable peers earned low returns. The raw high-minus-low return spread is 1.13% per month ($t = 5.52$). The corresponding alphas from standard factor models are also economically large and statistically significant. For example, the Fama-

[4]The average beta exceeds one across portfolios because portfolio characteristics are computed using equal weighting.

French five-factor alpha is 1.31% per month ($t = 5.91$), and the alpha after additionally controlling for momentum is 1.10% per month ($t = 4.84$).[5]

Panel B repeats the analysis in Panel A using alternative formation horizons by replacing the past one-month peer return with the past one-year or past one-week peer return. When RPSIM is constructed using past one-year peer returns, we continue to obtain a positive five-factor alpha of 0.62% per month ($t = 2.73$). However, the abnormal performance is largely absorbed once momentum is included (0.33% per month; $t = 0.67$). This pattern differentiates our results from the style-investing evidence in Wahal and Yavuz (2013), who emphasize the predictive ability of 12-month style returns, with weaker predictability at shorter horizons. Moreover, constructing RPSIM from past one-week peer returns yields an even stronger strategy performance than the baseline monthly specification.

Huang (2019) show that after investors realize positive returns on a stock, their propensity to purchase stocks in the same industry increases. To assess whether our price-based return predictability instead reflects this industry experience-induced trading or classic industry momentum (Moskowitz and Grinblatt, 1999), we form portfolios using the past-month average return of price-comparable peers drawn from different industries. Panel C of Table 2 reports the results. The price-based predictability remains economically large and statistically significant, with a five-factor alpha of 0.87% per month ($t = 4.43$). This evidence indicates that our findings are not fully explained by investors' increased tendency to buy within-industry stocks after positive experiences, nor by industry momentum. In the subsequent analysis, we also control for industry momentum more directly using Fama and MacBeth (1973) regressions.

Panel C also reports two additional robustness tests. First, when defining price-comparable peers, we exclude stocks with overlapping analyst coverage (Ali and Hirshleifer, 2020). The resulting RPSIM strategy continues to deliver a significant five-factor alpha of 1.16% per month ($t = 5.99$). Second, we conduct a placebo test by constructing RPSIM using randomly matched peers. As expected, this placebo measure exhibits no predictive power for focal stocks' subsequent returns.

Lastly, Panel D reports results based on the unadjusted peer-return measure, PSIM. The evidence remains qualitatively unchanged. For example, the value-weighted high-minus-low

[5]In Appendix Table A1, we evaluate the strategy performance under alternative factor-model adjustments and find that the price-based return predictability remains economically meaningful and statistically significant.

return spread is 0.92% per month ($t = 4.22$), with a five-factor alpha of 1.13% ($t = 4.52$). Overall, these results indicate that price-based return predictability does not depend on the orthogonalization procedure.

3.2 Double Sorting

To further ensure that our price-based return predictability is not completely driven by known anomalies documented in the literature, we conduct a series of standard double-sorted portfolio tests that control for these anomalies one at a time. At the beginning of each month, we divide all firms in our sample into five groups based on a lagged anomaly characteristic (e.g., book-to-market). Within each anomaly quintile, we then sort firms into decile portfolios based on RPSIM. We hold the resulting portfolios for one month and compute value-weighted excess returns. To summarize the price-based return predictability net of the anomaly characteristic, we average the RPSIM-decile returns across the characteristic quintiles.

Table 3 shows that the price-based return predictability remains statistically significant after controlling for each anomaly characteristic in turn, including prominent return predictors such as size, book-to-market, profitability, investment, idiosyncratic volatility, maximum daily return, skewness, coskewness, illiquidity, past returns over multiple horizons, and industry momentum. Overall, the evidence indicates that our price-based return predictability is distinct from these existing anomalies. Importantly, the effect persists after controlling for past-return measures across horizons and for industry momentum, indicating that it is not driven solely by the correlation between RPSIM and past-return-based predictors.

3.3 Fama-MacBeth Regressions

The preceding subsections report average portfolio returns from single-sort and double-sort exercises. Although the double-sorting approach is intuitive, it cannot flexibly control for multiple return determinants simultaneously, because dependent sorts on more than two or three characteristics quickly become infeasible given the finite cross-sectional sample. We therefore complement the portfolio evidence with predictive regressions that test whether the average return of price-based peers predicts focal-stock returns in a manner consistent with the portfolio results. This regression framework allows us to include a rich set of controls in a single specification, thereby assessing whether the documented price-based predictability

merely reflects comovement with other well-known anomalies.

Besides the standard anomaly controls, we are particularly concerned that our key variable, RPSIM, may proxy for traditional style-based returns. The style-investing literature argues that investors simplify portfolio choice by grouping securities into styles (e.g., size and value/growth) and allocating capital at the style level rather than at the individual-stock level (Barberis and Shleifer, 2003). Consistent with this mechanism, prior work documents positive-feedback trading in style portfolios (Teo and Woo, 2004; Froot and Teo, 2008), which can in turn generate return predictability in lagged style returns (Wahal and Yavuz, 2013). Since size and the book-to-market ratio are among the most widely used style dimensions in practice and are actively targeted by institutional and retail investors (Kumar, 2009; Froot and Teo, 2008), we follow Wahal and Yavuz (2013) and explicitly control for size-BM style returns.

Table 4 reports Fama and MacBeth (1973) regressions of one-month-ahead stock returns on lagged RPSIM and controls for style and industry returns. Column (1) presents the baseline specification with RPSIM only, which shows that a one-standard deviation increase in lagged RPSIM implies an increase in the focal stock's subsequent monthly return of 0.164% (t=5.83). Columns (2) to (4) add style and industry returns measured over the past 1 month, 6 months, and 12 months, respectively, and column (5) includes style and industry returns at all three horizons jointly.

Across all specifications, RPSIM remains positive and statistically significant with a stable magnitude. Thus, our price-based return predictability is not subsumed by the traditional style investing or industry momentum. In addition, consistent with Wahal and Yavuz (2013), style returns exhibit predictive power primarily at longer horizons, whereas shorter-horizon (i.e., one-month) style returns are weak when included in isolation. Industry returns display strong continuation at multiple horizons, with positive and generally significant coefficients, consistent with the industry momentum literature (Moskowitz and Grinblatt, 1999). Importantly, controlling for these industry returns does not attenuate the RPSIM coefficient. Thus, our focus on the predictive power of the average short-term return of price-comparable stocks also distinguishes this paper from the style investing literature.

Overall, the regression evidence corroborates our portfolio results. The price-based return predictability persists after accounting for conventional style investing and industry momentum, and it remains economically meaningful and statistically significant in the presence of these controls.

3.4 Price-Based Return Predictability over the Long Horizon

Analogous to other return predictability patterns such as momentum (Jegadeesh and Titman, 1993), our price-based return predictability has several potential explanations. More specifically, two prevailing channels exist: (1) continued overreaction after portfolio formation, and (2) initial underreaction at the portfolio construction date. In the continued overreaction channel, if some stocks experienced high returns in the past, investors might pay more attention to stocks with comparable price level, creating higher demand for these stocks.[6] As a result, the focal stock tends to experience a higher return due to the price pressure.

For the underreaction story, predictability emerges because of slowly diffused fundamental news. If similarly priced stocks share correlated fundamentals and investors overlook this information, then the focal stock tends to be undervalued (overvalued) when peer stocks have earned high (low) returns in the past. Abnormal returns result as focal stocks' prices converge to fundamental values. However, this case appears less plausible since the nominal price level itself provides little justification of firm fundamentals. To distinguish alternative channels underlying price-based return predictability, we begin by examining strategy returns over longer horizons.

Figure 1 presents price-based return predictability over long horizons. We form portfolios with long-term holding periods following the method of Jegadeesh and Titman (1993). The RPSIM strategy return increases gradually over the first 16 months and then begins to decline. The effect fully reverses and even becomes negative after about 30 months. Since the standard momentum also exhibits long-term reversal, we account for this potential confound by estimating six-factor alphas. The lower graph of Figure 1 shows that price-based predictability still diminishes and reverses over the long run. Controlling for the five-factor model yields the same conclusion. This long-run reversal pattern makes it less likely that our results reflect initial underreaction to fundamental news. The next section further explores the mechanism by providing evidence based on stock splits, investor trading, and investor attention.

[6]Indeed, Huang (2019) finds that after personally experiencing a positive rather than negative investment return from a stock, investors' propensity to buy stocks in the same industry or stocks with similar size and book-to-market significantly increases.

4 Inspecting the Mechanism

In this section, we examine the key mechanisms underlying the documented price-based return predictability. First, we use stock splits as a laboratory to study the joint role of price-based categorical thinking and experience effect. Second, we investigate the subsequent retail order imbalance to scrutinize the continued overreaction hypothesis. Lastly, we show that attention spills over to price-comparable stocks.

4.1 Evidence Based on Stock Splits

Our price-based predictability relies on the *interaction* of two elements: (1) categorical thinking by price (e.g., Green and Hwang (2009)), and (2) anchoring on past investment experience (e.g., Huang (2019)). In particular, either price-based categorical thinking or experience-driven positive feedback trading alone cannot generate our result. In this section, we use stock splits as shocks to nomimal prices to further inspect the mechanism underlying our result.

Specifically, to the extent that investors anchor to prior investment experience, they may still pay attention to the pre-split price level and track the corresponding set of stocks in their trading. For instance, Campbell et al. (2014) find that individuals repeat investment behaviors and invest in categories that performed well for them in the recent past. Kuhnen et al. (2017) show that investors' portfolio choices are sticky over time, as prior investment choices hinder belief updating. Bazley et al. (2021) find over reliance on past decisions in investors' behavior. Since we define price-comparable stocks using the most recent month-end price, peer stocks based on the post-split nominal price may diverge from investors' mentally associated benchmark through past experience. Therefore, if price-based return predictability is experience-driven, then we would observe reduced predictive ability of RPSIM after stock splits due to this mismatch between investors' reference set and the mechanically redefined peer group.

Stock splits also allow us to distinguish between price-based categorical thinking and alternative style-based investing (e.g., Barberis and Shleifer (2003)). We expect that the predictive ability of peer stocks' returns should be unaffected by split events when those peers are not defined along nominal price, e.g., when identified through industry (Moskowitz and Grinblatt, 1999) or other characteristic classifications (Wahal and Yavuz, 2013).

We collect split event data from CRSP and design our empirical analysis following approaches in related studies, such as Green and Hwang (2009), Birru and Wang (2016), Shue and Townsend (2021), and Du et al. (2025). In this section, we focus on 2-for-1 splits, the most prevalent split scheme. We further exploit a difference-in-differences design by comparing results from 2-for-1 splits and moderate-factor splits in later analysis.

4.1.1 Revisit Green and Hwang (2009) with predictive regressions

Before implementing our main split analysis, we perform a practice similar to the empirical design of Green and Hwang (2009). Specifically, we first replicate their comovement result by estimating the following regression for each stock split, before and after the event:

$$R_{i,d} = \alpha_i + \beta_{LowPrc,i} R_{LowPrc,i,d} + \beta_{HighPrc,i} R_{HighPrc,i,d} + \varepsilon_{i,d}, \tag{2}$$

where $R_{i,d}$ is the return of stock i on day d and $R_{LowPrc,i,d}$ ($R_{HighPrc,i,d}$) is stock i's respective low (high) price index on day d. We follow Green and Hwang (2009) in defining the low- and high-price groups and in constructing $R_{LowPrc,i,d}$ and $R_{HighPrc,i,d}$.[7] We then modify the design slightly by estimating predictive regressions:

$$R_{i,d+1} = \alpha_i + \beta_{LowPrc,i} R_{LowPrc,i,d} + \beta_{HighPrc,i} R_{HighPrc,i,d} + \beta_{Focal,i} R_{i,d} + \varepsilon_{i,d+1}, \tag{3}$$

where the main difference is that we replace contemporaneous return on the left-hand-side ($R_{i,d}$) with the one-day-ahead return ($R_{i,d+1}$) and control for the focal stock's lagged return to purge out the short-term reversal effect. For both specifications, the pre-event regression is run over the 12-month period ending one month before the split implementation, and the post-event regression is run over the 12-month period starting one month after the split implementation. The sample consists of 2-for-1 split events spanning 1964 to 2024. When summarizing the estimation results, standard errors are clustered by month when calculating t-statistics.

[7] Specifically, the low-price group (LowPrc) and the high-price group (HighPrc) are given by: $LowPrc \in \left[P_{post} - \frac{P_{pre}-P_{post}}{2}, P_{post} + \frac{P_{pre}-P_{post}}{2}\right]$, $HighPrc \in \left[P_{pre} - \frac{P_{pre}-P_{post}}{2}, P_{pre} + \frac{P_{pre}-P_{post}}{2}\right]$, where P_{pre} and P_{post} are the pre- and post-split stock prices measured one day before the split. As in Green and Hwang (2009), we focus on 2-for-1 splits, meaning $P_{pre} = 2P_{post}$. Therefore, the low-price group and the high-price group can also be written as $LowPrc \in [\frac{1}{4}P_{pre}, \frac{3}{4}P_{pre}]$ and $HighPrc \in [\frac{3}{4}P_{pre}, \frac{5}{4}P_{pre}]$. Then, $R_{LowPrc,i,d}$ ($R_{HighPrc,i,d}$) is the value-weighted average return of stocks belong to LowPrc (HighPrc), excluding the focal stock.

Table 5 reports the results. We confirm that the price-based comovement documented by Green and Hwang (2009) remains highly significant in our extended sample. After splits, stocks comove more (less) with low-priced (high-priced) stocks. Panel A shows that the cross-sectional mean change in $\beta_{LowPrc,i}$ is 0.206 (t=9.01), whereas the cross-sectional mean change in $\beta_{HighPrc,i}$ is -0.085 (t=-4.39). Panel B presents the predictive regression results. First, we find that price-comparable stocks' past return exhibit positive predictive ability in the Green and Hwang (2009) setting. Moreover, split event alter this predictive relation. After stock splits, the coefficient on the low-price index ($\beta_{LowPrc,i}$) in predictive regression (3) increases from 0.063 to 0.144, while the coefficient on the high-price index ($\beta_{HighPrc,i}$) declines from 0.168 to 0.117. The corresponding changes in average coefficients are also statistically significant, with the difference in the mean $\beta_{LowPrc,i}$ being 0.081 (t=3.68) and the difference in the mean $\beta_{HighPrc,i}$ being -0.051 (t=-2.35).

4.1.2 Attenuated price-based return predictability after stock splits

We now turn to the price-based return predictability documented in this paper. To capture variations in predictability, we define a dummy variable, PredictPower$_{i,t}$, for each stock-month. Specifically, PredictPower$_{i,t}$ is set to one if either of the following conditions is met: (1) stock i's current monthly return and its RPSIM at $t-1$ are both in the top quintile; or (2) stock i's current monthly return and its RPSIM at $t-1$ are both in the bottom quintile. For all other cases, PredictPower$_{i,t}$ takes a value of zero. In other words, the average PredictPower$_{i,t}$ captures the probability that stocks with high (low) RPSIM in the past will outperform (underperform) in the subsequent month.

Figure 2 illustrates the timeline of variable construction and empirical test. We estimate the following regression to exmaine the impact of stock splits on price-based predictability:

$$\text{PredictPower}_{i,t+1} = \alpha + \sum_{k=-6}^{6} \beta_k \mathbb{I}\{EventMonth_{i,t} = k\} + \tau_t + \eta_i + \varepsilon_{i,t+1}, \tag{4}$$

where the dummy variable $\mathbb{I}\{EventMonth_{i,t} = k\}$ equals to one if month t is k months before or after a split event for stock i. The split event month is characterized by $k = 1$. We include year-month fixed effects (τ_t) and firm fixed effects (η_i). The sample is limited to six months before and after a split.[8] Since the dependent variable represents intertemporal relationship

[8]To be included in the sample, for each stock split event, we also require that there are no other split events for the same stock occurring within a 6-month window before or after the event. In addition, monthly

between RPSIM in period k and return in period $k+1$, we use $k=-2$ as our estimation benchmark.[9] The coefficients β_k $(k=-6,...,-1,1,....,6)$ thus measure the difference in predictive power in event month k relative to the benchmark (pre-split) period. If a split event attenuates price-based return predictability, one would expect to see a significantly negative β_k for $k \geq 1$.

Figure 3 presents the regression results. Consistent with our hypothesis, the predictive power of RPSIM declines sharply following stock splits. The estimated coefficient (β_k) is small in magnitude during the pre-event window, while becoming significantly negative in the post-event period. This pattern suggests that price-based categorical thinking alone cannot fully explain our predictability findings. As Green and Hwang (2009) document, stocks exhibit increased (decreased) comovement with low-priced (high-priced) stocks after splits. Different from Green and Hwang (2009), who focus on *contemporaneous* and time-series return correlations, our results highlight the role of investor experience as a joint and critical determinant of *intertemporal* and cross-sectional return predictability.

4.1.3 Placebo tests based on alternative stock categories

To further examine the role of price-based categorical thinking and distinguish our story from other channels, we conduct a placebo test by re-estimating equation (4) using alternative return variables. Specifically, we consider two control variables employed in our Fama-MacBeth regressions, the one-month industry return (Moskowitz and Grinblatt, 1999) and one-month style return (Wahal and Yavuz, 2013). We also consider four alternative characteristic-based peer returns, including size, book-to-market ratio, profitability, and investment. For each characteristic, we follow Section 2.1 and define peer stocks analogously as when defining price-based peers, and then we calculate value-weighted average monthly return of the focal stocks' characteristic-based peer stocks.[10] We conjecture that split events

stock returns should be non-missing throughout this window. As in our main analysis, the sample is restricted to stocks with a share price of at least \$1 at the end of each month. After these filters, our initial sample consists of 4178 2-for-1 split events spanning from 1964 to 2024.

[9]Specifically, β_k captures the predictive ability of period k's RPSIM for return in $k+1$. For $k=-1$, since stocks would experience splits in the upcoming month ($k=1$), the estimated β_{-1} might be contaminated by many confounding factors such as shifts in comovement (Green and Hwang, 2009), changes in investor clientele (Dhar et al., 2004), or changes in liquidity (Copeland, 1979; Conroy et al., 1990; Easley et al., 2001; Lin et al., 2009). Therefore, β_{-2} serves as a more plausible benchmark since it is purely based on pre-split information. Nevertheless, we use $k=-1$ (one-month prior to split) for robustness check and find similar results.

[10]To align with our main specification, we use the orthogonal versions of these alternative peer return variables. Specifically, for each variable, we estimate monthly cross-sectional regressions of the variable on

should not affect the predictive ability of these non-price-based variables in the same way as our price-based predictability.

Figure 4 shows estimation results based on alternative definitions of peer stock. In sharp contrast, we detect no significant changes in return predictability around stock splits. The estimated coefficients appear randomly distributed and mostly insignificant throughout the event window. As such, characteristics-driven style investing (Barberis and Shleifer, 2003; Wahal and Yavuz, 2013) is unlikely to drive our priced-based return predictability. In addition, this pattern also suggests that potential changes in liquidity condition (Copeland, 1979; Conroy et al., 1990; Easley et al., 2001; Lin et al., 2009) around stock splits are less likely to be the main force underlying our result, since otherwise one should observe material shifts in predictive power of these characteristics-based predictors. These placebo results highlight the distinctive roles of price-based categorical thinking and investment experience in generating the predictable return pattern documented in this paper.

4.1.4 Comparing 2-for-1 and moderate-factor splits

A potential concern with our analysis is that unobservable fundamental factors might drive split decisions. We argue, however, that our findings still support the behavioral interpretation over other channels such as the sluggish information dissemination. First, even if splits partly reflect evolving fundamentals, such slow-moving forces are unlikely to generate an *immediate* decline in RPSIM's predictive power. Second, we employ a difference-in-differences design that compares 2-for-1 splits (a 50% price reduction) with a pool of moderate-factor schemes, specifically 3-for-2 splits (a 33% price reduction) and 5-for-4 splits (a 20% price reduction), which enables us to control for common fundamental factors underlying split decisions (Shue and Townsend, 2021).[11] We show that price-based return predictability weakens more following larger nominal price shocks (i.e., 2-for-1 splits), as the wedge between the post-split price level and prior investment experience becomes more substantial.

To examine the difference in the impact between 2-for-1 and moderate-factor splits, we

the focal stocks' monthly return and take the residuals. Our conclusions remain valid using the original peer return variables.

[11] Using the same construction procedure, the sample includes 4178 2-for-1 splits, 2688 3-for-2 splits, and 655 5-for-4 splits from 1964 to 2024.

estimate the following regression:

$$\begin{aligned}\text{PredictPower}_{i,t+1} = &\sum_{k=-6}^{6} \beta_k \mathbb{I}\{EventMonth_{i,t} = k\} \\ &+ \sum_{k=-6}^{6} \gamma_k \mathbb{I}\{EventMonth_{i,t} = k\} \times \mathbb{I}\{TwoForOne_{i,t}\} \\ &+ \mathbb{I}\{TwoForOne_{i,t}\} + \alpha + \tau_t + \eta_i + \varepsilon_{i,t+1},\end{aligned} \tag{5}$$

where $\mathbb{I}\{TwoForOne_{i,t}\}$ is an indicator variable that is equal to one if the split event for stock i in month t is a 2-for-1 split, and zero if it is a 3-for-2 split or 5-for-4 split. Similar to equation (4), we use two-month prior to the split event (i.e., $k = -2$) as our regression benchmark. We expect that the coefficient on the interaction term, γ_k, should be significantly negative in the post-event period ($k \geq 1$).

Figure 5 presents the regression results. We find support for the parallel trends assumption, as the estimated γ_k is insignificant for $k < 0$. More importantly, γ_k becomes negative and statistically significant after stock splits, indicating that the price-based return predictability diminishes more substantially following 2-for-1 splits. Since changes in fundamentals are unlikely to happen precisely in split execution months, and equation (5) controls for common fundamental movements, our result suggests that the documented variation in RPSIM's predictive ability is mainly explained by nominal price-induced disruption in investor perceptions.

We also re-estimate regression (5) using alternative characteristics-based variables as placebo predictors. Figure 6 presents the results. Consistent with our interpretation and previous results, we find no material change in predictive power around stock splits when the predictor is not constructed from nominal share price.

4.1.5 Placebo tests based on alternative corporate events

To complement our analysis, we examine a series of major corporate events that also potentially affect stock prices. Specifically, we employ 23 event categories related to firms' real economic activities, as studied by Kwon and Tang (2025). Quarterly earnings announcement dates are obtained from Compustat, while the dates for the remaining 22 event types (e.g., board changes and annual meetings) are extracted from the Capital IQ

Key Developments database.[12] For each event included in our analysis, we require that no stock split events occur within a six-month period preceding or following the event. We expect that these fundamental-driven corporate events exert minimal impact on price-based return predictability.

In addition, a large literature shows that demand shocks from index reconstitutions lead to abnormal returns and changes in stock return comovement (Harris and Gurel, 1986; Shleifer, 1986; Barberis et al., 2005; Chang et al., 2015). Since reconstitutions typically have requirements on market capitalization (e.g., Russell indexes) and fundamental performance (e.g., S&P 500 index), and not targeting on nominal prices, we expect that they would also have little influence on the predictive ability of RPSIM. Data on firm-level index reconstitution events are obtained from Greenwood and Sammon (2025).[13] We define the event month as the month in which the index-constituent change is implemented and apply the same filters used for other corporate events.

Figure 7 presents coefficients from equation (4), in which we define the event window using each event category. The final subfigure shows the result for index reconstitutions. Overall, these corporate events fail to produce the pattern observed in our benchmark test (Figure 3). In most cases, RPSIM's predictive power is statistically insignificant both before and after the event, or it changes in a direction inconsistent with our mechanism. These placebo results therefore reinforce our interpretation that price-based return predictability arises primarily from price-based categorization interacting with the experience effect, rather than from alternative channels such as sluggish diffusion of fundamental news.

Overall, stock splits provide a useful setting to inspect the key channel underlying our asset pricing results. Having established the split-based evidence on predictability, we next test two additional implications of our hypothesis for investor trading and investor attention.

4.2 Evidence Based on Order Imbalance

As noted in Section 3.4, both continued overreaction and initial underreaction could be responsible for price-based return predictability. Section 3.4 shows that such predictability

[12]The original sample of Kwon and Tang (2025) comprised 24 corporate event categories. However, structural changes in the Capital IQ Key Developments database in December 2023 resulted in the removal of Credit Watch-related events. Therefore, we focus on the remaining 23 categories.

[13]These events consist of S&P 500 additions and deletions, as well as additions to the Russell 1000 and 2000 from outside the Russell 3000 universe. We thank the authors for making the data available.

exhibits evident long-term reversal, whereas Section 4.1 suggests that fundamental factors play minimal role in driving declines in price-based return predictability after split events. To further tease out the channel, we examine trading metrics in this section. In particular, categorical thinking by price suggests that investors employing such a heuristic would display trading habitats among price-comparable stocks. As prior investment experience shapes current tastes, our mechanism predicts that, after price-comparable stocks realized high (low) returns recently, the effective demand for focal stocks should also increase (decrease) subsequently. Since retail investors are more likely to be drawn to categorical thinking and susceptible to experience anchoring, the demand response should be more evident for retail trading.

We study subsequent order imbalance for RPSIM portfolios following the portfolio formation period. Retail trading activity is proxied by small orders and calculated based on the Lee and Ready (1991) algorithm.[14] We examine how order imbalance evolves in the month following portfolio formations. Figure 8 reports the result. It shows that the demand from retail investors increases more (decreases less) for stocks with high price-comparable stock returns in the past month than for stocks with low price-comparable stock returns. This higher subsequent demand for stocks with high price-comparable stock returns could be partially responsible for the higher subsequent returns for these stocks.

Notice that the order imbalance pattern alone does not rule out the initial underreaction hypothesis. In particular, if firms with high RPSIM are underpriced more compared with firms with low RPSIM, then investors could be rationally trading on such mispricing. If this is the case, however, one should expect a stronger response for large orders, where institutional investors dominate. Figure 8 shows that the pattern based on large, potentially institutional, order imbalance is minimal. Thus, the stronger subsequent demand from retail investors for high price-comparable stock returns suggests that this enhanced retail trading is potentially more consistent with the continued overreaction channel.

Beyond retail trading, we also examine trading volume. Short-sale constraints imply that price-based return predictability is driven primarily by buy-side pressure. As such, stocks whose price-comparable peers earned high returns in the past should exhibit abnormally elevated trading volume. Appendix Table A2 confirms our hypothesis. We find that higher RPSIM is indeed associated with higher abnormal turnover in the next month.

[14]The calculation details are provided in Appendix A.

4.3 Evidence Based on Retail Investor Attention

When investors anchor to experience and employ categorical thinking, recent performance of same-category stocks attracts attention and influence trading decisions. In Section 4.2, we have shown that similarly priced stocks' recent returns indeed triggers retail investments. Here we examine response in attention. As claimed in the introduction, if investors think along price, then after experiencing a good return on stock investment, they may pay more attention to stocks with comparable price level. This attention allocation translates into purchase behavior and leads to continued overreaction to recent performance. Similarly, investors are less willing to hold a stock after experiencing unfavorable performance in investment of similarly priced stocks. This argument posits that the magnitude of price-comparable stocks' recent returns should be positively associated with retail investors' attention to the focal stock.

We measure retail investor attention using abnormal Google search volume (Da et al., 2011), defined as the log difference between the current Google search volume index and the medium index in the past three months. Table 6 examines the relationship between RPSIM and subsequent retail attention through panel regressions. First, we consider the magnitude of price-based peers' recent returns in Panel A:

$$\text{Retail Attention}_{i,t+1} = \alpha + \beta|RPSIM_{i,t}| + Controls_{i,t} + \varepsilon_{i,t+1}. \tag{6}$$

Indeed, we find that the magnitude of RPSIM positively and significantly predicts retail investor attention in the next month. For example, column (3) shows that an increase in the magnitude of lagged RPSIM by 1% indicates an increase in future abnormal Google search volume of around 0.198 (t=3.55). Since Da et al. (2011) documents that retail investor attention is highly persistent, we further include lagged retail attention as an additional control in columns (4) and (5). We find that the estimated coefficient on lagged |RPSIM| remains positive and highly significant.

In Panel B of Table 6, we separate positive and negative parts of price-based peers' recent returns by estimating the following regression:

$$\text{Retail Attention}_{i,t+1} = \alpha + \beta_1 RPSIM^{+}_{i,t} + \beta_2 RPSIM^{-}_{i,t} + Controls_{i,t} + \varepsilon_{i,t+1}, \tag{7}$$

where $RPSIM^{+}_{i,t} = \max\{RPSIM^{+}_{i,t}, 0\}$ and $RPSIM^{-}_{i,t} = \min\{RPSIM^{+}_{i,t}, 0\}$. Intuitively, if attention spills over to similarly priced stocks after a high recent performance, then one

would expect a positive coefficient on $RPSIM^{+}_{i,t}$. If unfavorable experience also triggers attention (e.g., investors are less likely to hold a stock with similar price level), then one should expect a negative coefficient on $RPSIM^{-}_{i,t}$. Regressions in Panel B confirm our story. After controlling for lagged retail investor attention, column (6) shows that an increase in the positive part of RPSIM by 1% implies increases in retail investors' attention to the focal stock by 0.215 (t=2.66); a decrease in the negative part of RPSIM by 1%, similarly, suggests an increase in retail attention by 0.162 (t=-2.81).

5 Further Robustness Checks

This section provides heterogeneity analysis based on sample period, firm size, idiosyncratic volatility, and institutional ownership. We also extend the analysis to show that price-based return predictability is an international phenomenon.

5.1 Subsample Analysis

Panel A of Table 7 reports a subsample analysis in which we split the full sample into an early and a recent period. The price-based return predictability is stronger in the early subsample: the monthly Fama-French three-factor alpha equals 1.35% (t = 7.19) in the early period and 0.85% (t = 3.88) in the recent period. This pattern is consistent with the view that the mechanism underlying price-based predictability operates primarily through retail trading. To the extent that direct retail participation in individual stocks declines in the more recent period, the resulting attenuation in retail-driven demand spillovers would weaken the predictability pattern. This time-series pattern contrasts with the evidence on traditional style investing. Wahal and Yavuz (2013) show that style returns predict individual-stock returns in the cross-section primarily in the post-1988 period, whereas the predictive power of past style returns is largely indistinguishable from zero earlier. They attribute the stronger predictability in the later period to the increased use of size and value categorization in mutual funds and institutional portfolios.

A potential reconciliation is that the ultimate investors in mutual funds are predominantly retail investors. Thus, even if style-based predictability in the later period reflects style allocation by delegated asset management, it remains consistent with a retail origin of style-level flows. Under this interpretation, greater retail participation through style-oriented

mutual funds could amplify the predictive power of style returns in the recent period, while reduced direct retail participation in individual stocks could dampen the predictive power of similarly priced peers' returns.

Panel B of Table 7 reports subsamples by firm size. We classify firms as large if their market capitalization exceeds the NYSE median and as small otherwise. The price-based return predictability remains statistically significant in both groups. For example, the high-minus-low RPSIM portfolio among large firms continues to deliver a monthly three-factor alpha of 0.38% ($t = 4.06$), although the magnitude is smaller than among small firms. This evidence is informative since Wahal and Yavuz (2013) find little cross-sectional predictive power of past style returns among large stocks, suggesting that value-growth style returns alone do not account for return variation among large stocks.

We also split the sample into two groups based on idiosyncratic volatility in Panel C. The price-based predictability is again present in both subsamples. As expected, the effect is weaker among low-IVol firms but remains economically meaningful and statistically significant, with a monthly three-factor alpha of 0.52% (t =5.03). Overall, the evidence indicates that price-based return predictability is not confined to small or difficult-to-arbitrage stocks, although it is stronger in those segments, consistent with limits-to-arbitrage slowing the correction of mispricing. At the same time, the persistence of the effect among large and low-IVol firms suggests that arbitrage activity does not fully eliminate the demand spillovers induced by the categorical thinking-experience effect interaction, potentially because such bias is not immediately salient to sophisticated traders.

If retail investors are more prone to categorical thinking and susceptible to the experience effect, the price-based return predictability should be stronger among firms with lower institutional ownership. Panel D of Table 7 reports average returns for portfolios formed by sequential sorts on institutional ownership and RPSIM. The high-institutional-ownership (high-IO) and low-institutional-ownership (low-IO) subsamples comprise the top and bottom halves of stocks, respectively, based on size-adjusted institutional ownership constructed following Nagel (2005).

The three-factor alpha of the RPSIM strategy is 1.10% per month among low-IO firms and 0.85% per month among high-IO firms. The same qualitative pattern holds for the raw excess return and six-factor alphas. Using mutual-fund holdings and institutional trading data, Teo and Woo (2004) and Froot and Teo (2008) document style-level trading by institutional investors. In contrast to the predictive power of style returns, which is often

attributed to institutional trading activity[15], our evidence points to a primarily retail-driven mechanism. The price-based return predictability is substantially larger among firms with lower institutional ownership.

5.2 Evidence from International Markets

In this subsection, we use international data to assess whether the patterns documented in the U.S. generalize across markets. Since trading intensity varies substantially across markets, a fixed calendar horizon may correspond to different volume-time horizons. For example, in high-turnover markets, one week of trading can be comparable (in terms of trading activity) to a month in low-turnover markets. In extremely high-turnover markets (e.g., China), investors may therefore respond more strongly to weekly, rather than monthly, past returns. Motivated by this heterogeneity, we partition markets into two groups based on aggregate turnover, defined as the average annual turnover for each country/area across all available years. For markets with aggregate turnover above the median, we form portfolios using the past one-week average residual return of price-comparable stocks and hold portfolios for one week. For markets with aggregate turnover below the median, we continue to use monthly RPSIM as the predictor and hold portfolios for one month. More broadly, aggregate turnover may proxy for retail participation, as individual investors tend to trade more actively.

We implement this procedure for the 38 markets reported in Table 8. Given limited cross-sectional coverage in some markets, we compute RPSIM using the average returns of the 30 most similarly priced peers. In the high-turnover group, 16 of 19 markets exhibit a positive price-based predictability, and 11 are statistically significant. Because the international sample is shorter than the U.S. sample and has fewer cross-sectional observations, statistical power at the individual-market level is generally attenuated, as expected. Nevertheless, the average spread in the high-turnover group is economically large, with an annualized return spread of 7.85% ($t = 7.71$). Moreover, only two of the 19 high-turnover markets display negative spreads, and none is statistically significant. By contrast, the low-turnover group exhibits a smaller and statistically insignificant average spread, with an annualized return spread of 2.89% ($t = 2.95$). Balasubramaniam et al. (2023) document strong price-related investor clienteles in India, while Liu and Shi (2021) show that Chinese retail investors substantially segment their choices by stock price. Consistent with these findings, Table 8

[15]Ultimately, individuals are the end investors in style mutual funds and may therefore contribute to the predictability of style returns. Nonetheless, mutual fund managers retain discretion over stock selection.

shows that both India (IND) and Mainland China (CHN) markets exhibit significant price-based return predictability. Overall, these results are consistent with the interpretation that greater retail participation amplifies price-based return predictability.

6 Conclusion

In this study, we investigate the asset pricing implications of the joint effect of price-based categorical thinking and the experience effect. We find that firms with high price-comparable stock returns tend to earn significantly higher future returns than firms with low price-comparable stock returns. We argue that this finding is consistent with the view that investors' demand for firms with high similarly priced stock returns increases because of price-based style investing behavior. We also show that our price-based return predictability is not completely driven by traditional size and value style investing and other well-known anomalies. Moreover, price-based return predictability is weakened after stock splits, leading clean evidence on price-based categorical thinking.

References

Ali, U., and D. Hirshleifer. 2020. Shared analyst coverage: Unifying momentum spillover effects. *Journal of Financial Economics* 136:649–675.

Amihud, Y. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets* 5:31–56.

Anagol, S., V. Balasubramaniam, and T. Ramadorai. 2021. Learning from noise: Evidence from India's IPO lotteries. *Journal of Financial Economics* 140:965–986.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *The journal of finance* 61:259–299.

Balasubramaniam, V., J. Y. Campbell, T. Ramadorai, and B. Ranish. 2023. Who owns what? A factor model for direct stockholding. *The Journal of Finance* 78:1545–1591.

Bali, T. G., N. Cakici, and R. F. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of financial economics* 99:427–446.

Barber, B. M., S. Lin, and T. Odean. 2019. Mediating Investor Attention. *Working paper* .

Barber, B. M., and T. Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies* 21:785–818.

Barber, B. M., T. Odean, and N. Zhu. 2008. Do retail trades move markets? *The Review of Financial Studies* 22:151–186.

Barberis, N., and A. Shleifer. 2003. Style investing. *Journal of financial Economics* 68:161–199.

Barberis, N., A. Shleifer, and J. Wurgler. 2005. Comovement. *Journal of financial economics* 75:283–317.

Bazley, W., G. M. Korniotis, and G. R. Samanez-Larkin. 2021. Can the Past Hinder Investor Learning? *Available at SSRN 2846504* .

Ben-David, I., J. Birru, and V. Prokopenya. 2018. Uninformative feedback and risk taking: Evidence from retail forex trading. *Review of Finance* 22:2009–2036.

Birru, J., and B. Wang. 2016. Nominal price illusion. *Journal of Financial Economics* 119:578–598.

Bordalo, P., G. Burro, K. Coffman, N. Gennaioli, and A. Shleifer. 2025a. Imagining the future: memory, simulation, and beliefs. *Review of Economic Studies* 92:1532–1563.

Bordalo, P., J. Conlon, N. Gennaioli, S. Kwon, and A. Shleifer. 2025b. How people use statistics. *Review of Economic Studies* p. rdaf022.

Campbell, J. Y., J. Hilscher, and J. Szilagyi. 2008. In search of distress risk. *The Journal of finance* 63:2899–2939.

Campbell, J. Y., T. Ramadorai, and B. Ranish. 2014. Getting better or feeling better? How equity investors respond to investment experience. Tech. rep., National Bureau of Economic Research.

Carhart, M. M. 1997. On persistence in mutual fund performance. *The Journal of finance* 52:57–82.

Chang, Y.-C., H. Hong, and I. Liskovich. 2015. Regression discontinuity and the price effects of stock market indexing. *The Review of Financial Studies* 28:212–246.

Chen, X., L. An, Z. Wang, and J. Yu. 2023. Attention Spillover in Asset Pricing. *The Journal of Finance* 78:3515–3559.

Conroy, R. M., R. S. Harris, and B. A. Benet. 1990. The effects of stock splits on bid-ask spreads. *The Journal of Finance* 45:1285–1295.

Copeland, T. E. 1979. Liquidity changes following stock splits. *The Journal of Finance* 34:115–141.

Da, Z., J. Engelberg, and P. Gao. 2011. In search of attention. *The journal of finance* 66:1461–1499.

Daniel, K., and D. Hirshleifer. 2015. Overconfident investors, predictable returns, and excessive trading. *Journal of Economic Perspectives* 29:61–88.

Daniel, K., D. Hirshleifer, and L. Sun. 2020. Short-and long-horizon behavioral factors. *The review of financial studies* 33:1673–1736.

Dhar, R., W. N. Goetzmann, and N. Zhu. 2004. The impact of clientele changes: Evidence from stock splits. *Yale ICF working paper* .

Du, J., D. Huang, Y.-J. Liu, Y. Shi, A. Subrahmanyam, and H. Zhang. 2025. Nominal prices, retail investor participation, and return momentum. *Management Science* .

Easley, D., M. O'hara, and G. Saar. 2001. How stock splits affect trading: A microstructure approach. *Journal of Financial and Quantitative analysis* 36:25–51.

Fama, E. F., and K. R. French. 1996. Multifactor explanations of asset pricing anomalies. *The journal of finance* 51:55–84.

Fama, E. F., and K. R. French. 2015. A five-factor asset pricing model. *Journal of financial economics* 116:1–22.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of political economy* 81:607–636.

Froot, K., and M. Teo. 2008. Style investing and institutional investors. *Journal of Financial and Quantitative Analysis* 43:883–906.

Gabaix, X. 2019. Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1*, vol. 2, pp. 261–343. Elsevier.

Gao, H., D. Shi, and B. Zhao. 2021. Does good luck make people overconfident? Evidence from a natural experiment in the stock market. *Journal of Corporate Finance* 68:101933.

Gao, P., C. A. Parsons, and J. Shen. 2018. Global relation between financial distress and equity returns. *The Review of Financial Studies* 31:239–277.

Gompers, P. A., and A. Metrick. 2001. Institutional investors and equity prices. *The quarterly journal of Economics* 116:229–259.

Green, T. C., and B.-H. Hwang. 2009. Price-based return comovement. *Journal of Financial Economics* 93:37–50.

Greenwood, R., and M. Sammon. 2025. The disappearing index effect. *The Journal of Finance*

80:657–698.

Harris, L., and E. Gurel. 1986. Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures. *the Journal of Finance* 41:815–829.

Harvey, C. R., and A. Siddique. 2000. Conditional skewness in asset pricing tests. *The Journal of finance* 55:1263–1295.

Hertwig, R., G. Barron, E. U. Weber, and I. Erev. 2004. Decisions from experience and the effect of rare events in risky choice. *Psychological science* 15:534–539.

Hou, K., G. A. Karolyi, and B.-C. Kho. 2011. What factors drive global stock returns? *The Review of Financial Studies* 24:2527–2574.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *The Review of Financial Studies* 28:650–705.

Huang, X. 2019. Mark Twain's Cat: Investment experience, categorical thinking, and stock selection. *Journal of Financial Economics* 131:404–432.

Hvidkjaer, S. 2006. A trade-based analysis of momentum. *The Review of Financial Studies* 19:457–491.

Ince, O. S., and R. B. Porter. 2006. Individual equity return data from Thomson Datastream: Handle with care! *Journal of Financial Research* 29:463–479.

Jegadeesh, N. 1990. Evidence of predictable behavior of security returns. *The Journal of finance* 45:881–898.

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

Kuhnen, C. M., S. Rudorf, and B. Weber. 2017. The effect of prior choices on expectations and subsequent portfolio decisions. Tech. rep., National Bureau of Economic Research.

Kumar, A. 2009. Dynamic style preferences of individual investors and stock returns. *Journal of Financial and Quantitative Analysis* 44:607–640.

Kumar, A., and C. M. Lee. 2006. Retail investor sentiment and return comovements. *The Journal of Finance* 61:2451–2486.

Kwon, S. Y., and J. Tang. 2025. Extreme categories and overreaction to news. *Review of Economic Studies* p. rdaf037.

Lee, C. M. 1992. Earnings news and small traders: An intraday analysis. *Journal of Accounting and Economics* 15:265–302.

Lee, C. M., and M. J. Ready. 1991. Inferring trade direction from intraday data. *The Journal of Finance* 46:733–746.

Lehmann, B. N. 1990. Fads, martingales, and market efficiency. *The Quarterly Journal of Economics* 105:1–28.

Lewellen, J. 2002. Momentum and autocorrelation in stock returns. *The Review of Financial Studies* 15:533–564.

Lin, J.-C., A. K. Singh, and W. Yu. 2009. Stock splits, trading continuity, and the cost of equity capital. *Journal of Financial Economics* 93:474–489.

Liu, Y.-J., and Y. Shi. 2021. Styles of Investing as Ideologies: Machine Learning Approach. *Working Paper* .

Lo, A. W., and A. C. MacKinlay. 1990. When are contrarian profits due to stock market overreaction? *The review of financial studies* 3:175–205.

Moskowitz, T. J., and M. Grinblatt. 1999. Do industries explain momentum? *The Journal of finance* 54:1249–1290.

Nagel, S. 2005. Short sales, institutional investors and the cross-section of stock returns. *Journal of financial economics* 78:277–309.

Pearson, N. D., Z. Yang, and Q. Zhang. 2021. The Chinese warrants bubble: Evidence from brokerage account records. *The Review of Financial Studies* 34:264–312.

Peng, L., and W. Xiong. 2006. Investor attention, overconfidence and category learning. *Journal of Financial Economics* 80:563–602.

Shleifer, A. 1986. Do demand curves for stocks slope down? *The Journal of Finance* 41:579–590.

Shue, K., and R. R. Townsend. 2021. Can the market multiply and divide? Non-proportional thinking in financial markets. *The Journal of Finance* 76:2307–2357.

Stambaugh, R. F., and Y. Yuan. 2017. Mispricing factors. *The review of financial studies* 30:1270–1315.

Teo, M., and S.-J. Woo. 2004. Style effects in the cross-section of stock returns. *Journal of Financial Economics* 74:367–398.

Wachter, J. A., and M. J. Kahana. 2024. A retrieved-context theory of financial decisions. *The Quarterly Journal of Economics* 139:1095–1147.

Wahal, S., and M. D. Yavuz. 2013. Style investing, comovement and return predictability. *Journal of Financial Economics* 107:136–154.

Weber, E. U., U. Böckenholt, D. J. Hilton, and B. Wallace. 1993. Determinants of diagnostic hypothesis generation: effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:1151.

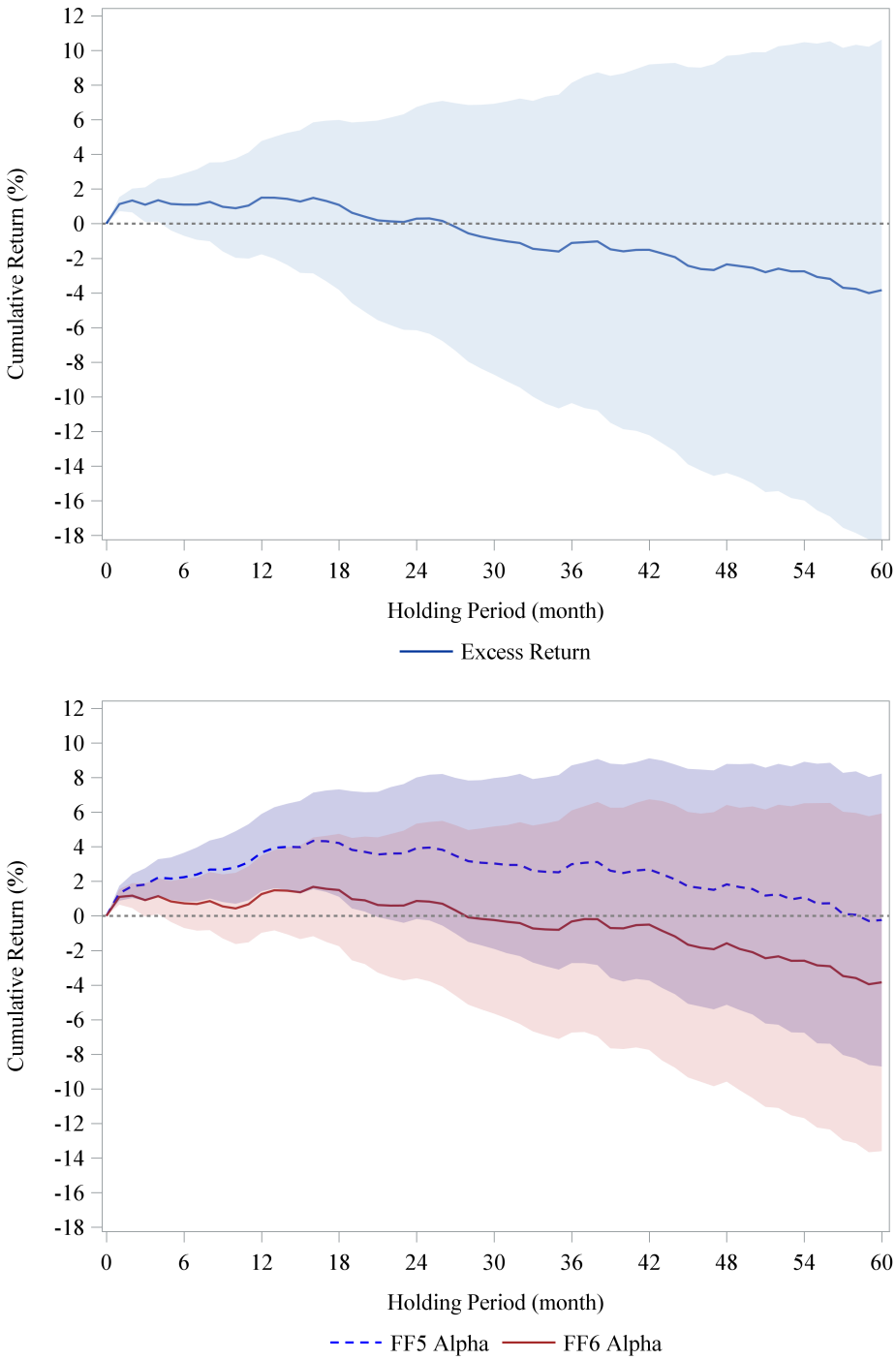


Figure 1: Price-Based Return Predictability over the Long-Run

This figure plots the price-based return predictability over long horizons using the method of Jegadeesh and Titman (1993). For strategy with a holding period of K months, each RPSIM decile consists of K subgroups, which are separately initiated in the previous K months. Portfolios are formed that equally weigh the subgroups. Then, the long-short strategy buys the top decile portfolio and sells the bottom decile portfolio. The lower graph reports abnormal returns adjusted by the Fama and French (2015) five-factor model and the six-factor model augmented by momentum. The sample period is from July 1963 to December 2024.

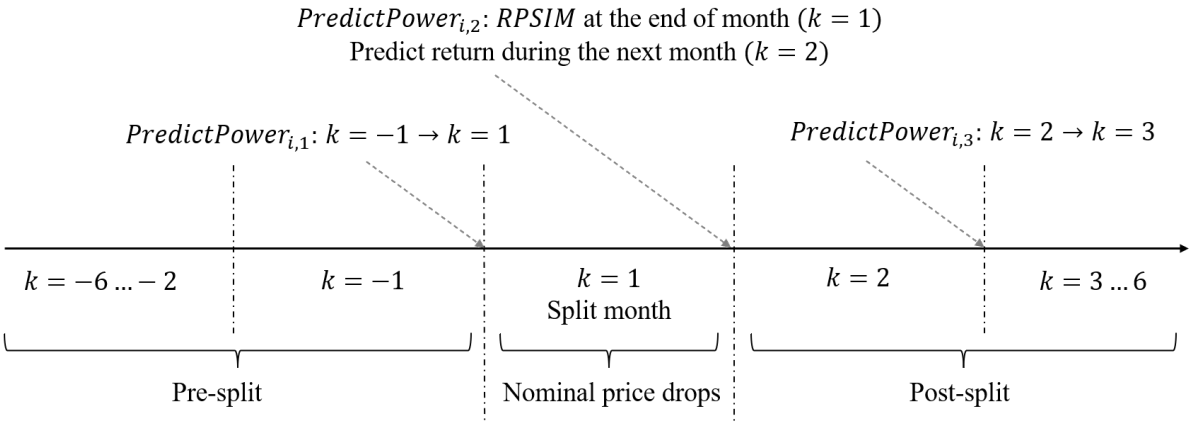


Figure 2: Timeline of Split Tests

This figure shows the timeline of stock split-related tests. We consider an event window spanning six months before and six months after a split. We define dummy variable $\mathbb{I}\{EventMonth_{i,t} = k\}$ for $k = -6, \ldots, -1, 1, \ldots, 6$, which equals one if month t is k months relative to the split event month. The month in which the split occurs is labeled $k = 1$. The main dependent variable of interest is $\text{PredictPower}_{i,t+1}$, the predictive power of month-t RPSIM for returns in month $t+1$. We compute RPSIM at the end of each month. Thus, for $k \geq 1$, RPSIM is calculated using month-end post-split prices, whereas for $k \leq -1$, RPSIM is calculated using month-end pre-split prices.

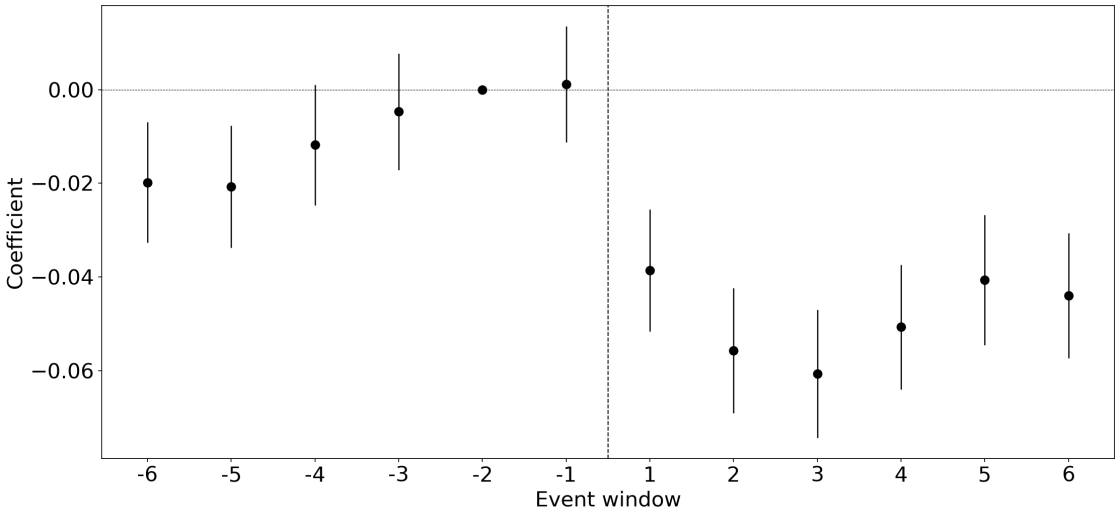


Figure 3: Price-Based Return Predictability and Stock Splits

This figure shows changes in the price-based predictability around stock splits by presenting the regression coefficients (β_k) from equation (4):

$$\text{PredictPower}_{i,t+1} = \alpha + \sum_{k=-6}^{6} \beta_k \mathbb{I}\{EventMonth_{i,t} = k\} + \tau_t + \eta_i + \varepsilon_{i,t+1},$$

where $k = -1$ serves as the estimation benchmark. The dependent variable $\text{PredictPower}_{i,t+1}$ is an indicator variable that equals one if both stock i's current monthly return $(t+1)$ and its lagged RPSIM (t) fall within either the top or bottom quintile, and zero otherwise. Dummy variable $\mathbb{I}\{EventMonth_{i,t} = k\}$ equals to one if month t is k months before or after a split event for stock i, with $k = 1$ denoting the split event month. The regression includes year-month fixed effects (τ_t) and firm fixed effects (η_i). Standard errors are clustered by stock and month. The sample period is from 1964 to 2024.

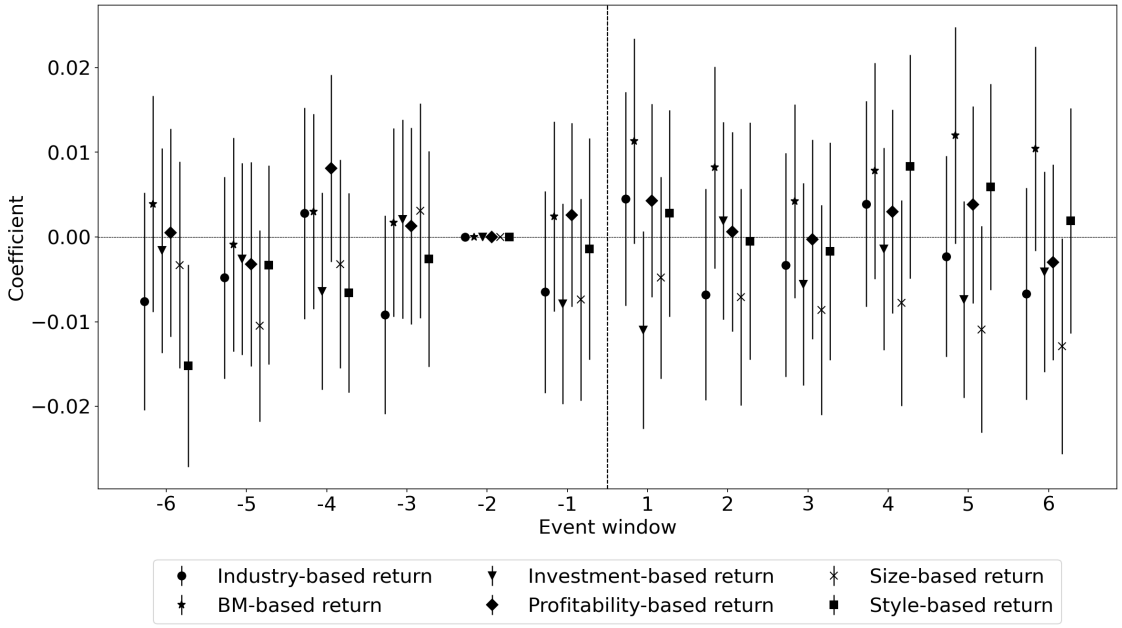


Figure 4: Splits and Predictability Based on Alternative Characteristics

This figure presents changes in alternative predictability pattern around stock splits. We consider six specifications: industry-based return (Moskowitz and Grinblatt, 1999), style-based return (Wahal and Yavuz, 2013), and four characteristic-based returns (size, book-to-market ratio, profitability, and investment). For each characteristic, we define peer stocks analogously as when defining price-based peers, then we calculate value-weighted average monthly return of the focal stocks' characteristic-based peer stocks. For each specification, the figure shows coefficients (β_k) from regression (4):

$$\text{PredictPower}_{i,t+1} = \alpha + \sum_{k=-6}^{6} \beta_k \mathbb{I}\{EventMonth_{i,t} = k\} + \tau_t + \eta_i + \varepsilon_{i,t+1},$$

where $k = -1$ serves as the estimation benchmark. The dependent variable $\text{PredictPower}_{i,t+1}$ is an indicator variable that equals one if both stock i's current monthly return $(t+1)$ and its lagged return-based predictor value (t) fall within either the top or bottom quintile, and zero otherwise. Dummy variable $\mathbb{I}\{EventMonth_{i,t} = k\}$ equals to one if month t is k months before or after a split event for stock i, with $k = 1$ denoting the split event month. Regressions include year-month fixed effects (τ_t) and firm fixed effects (η_i). Standard errors are clustered by stock and month. The sample period is from 1964 to 2024.

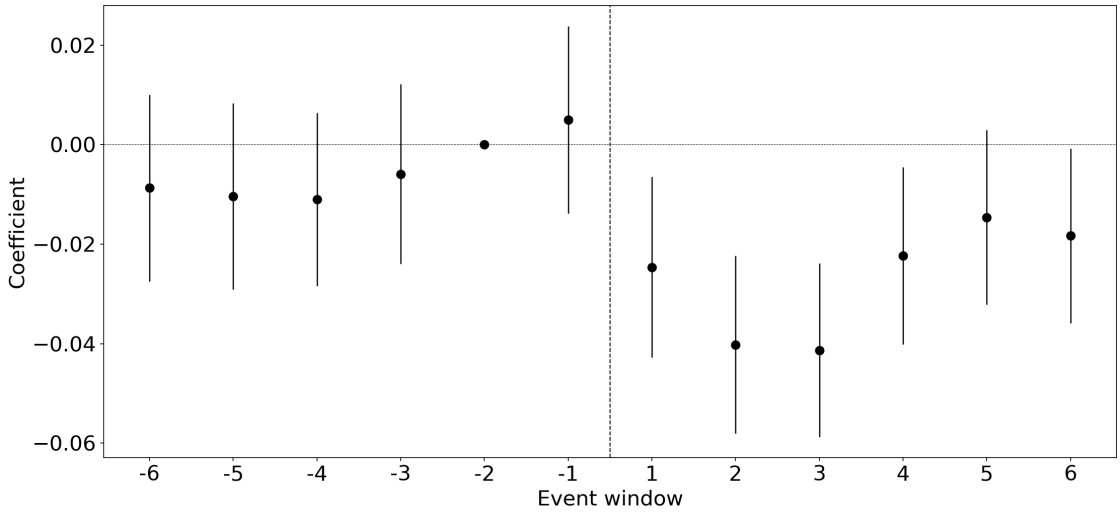


Figure 5: Comparison Between 2-for-1 and Moderate-Factor Splits

This figure presents the difference between 2-for-1 and moderate-factor (3-for-2 and 5-for-4) splits by showing regression coefficients (γ_k) from regression (5):

$$\begin{aligned}\text{PredictPower}_{i,t+1} = &\sum_{k=-6}^{6} \beta_k \mathbb{I}\{EventMonth_{i,t} = k\} \\ &+ \sum_{k=-6}^{6} \gamma_k \mathbb{I}\{EventMonth_{i,t} = k\} \times \mathbb{I}\{TwoForOne_{i,t}\} \\ &+ \mathbb{I}\{TwoForOne_{i,t}\} + \alpha + \tau_t + \eta_i + \varepsilon_{i,t+1},\end{aligned}$$

where $k = -1$ serves as the estimation benchmark. The dependent variable $\text{PredictPower}_{i,t+1}$ is an indicator variable that equals one if both stock i's current monthly return $(t+1)$ and its lagged RPSIM (t) fall within either the top or bottom quintile, and zero otherwise. Dummy variable $\mathbb{I}\{EventMonth_{i,t} = k\}$ equals to one if month t is k months before or after a split event for stock i, with $k = 1$ denoting the split event month. $\mathbb{I}\{TwoForOne_{i,t}\}$ is an indicator variable that takes a value of one if the split event for stock i in month t is a 2-for-1 split, and zero if it is a 3-for-2 split or 5-for-4 split. The regression includes year-month fixed effects (τ_t) and firm fixed effects (η_i). Standard errors are clustered by stock and month. The sample period is from 1964 to 2024.

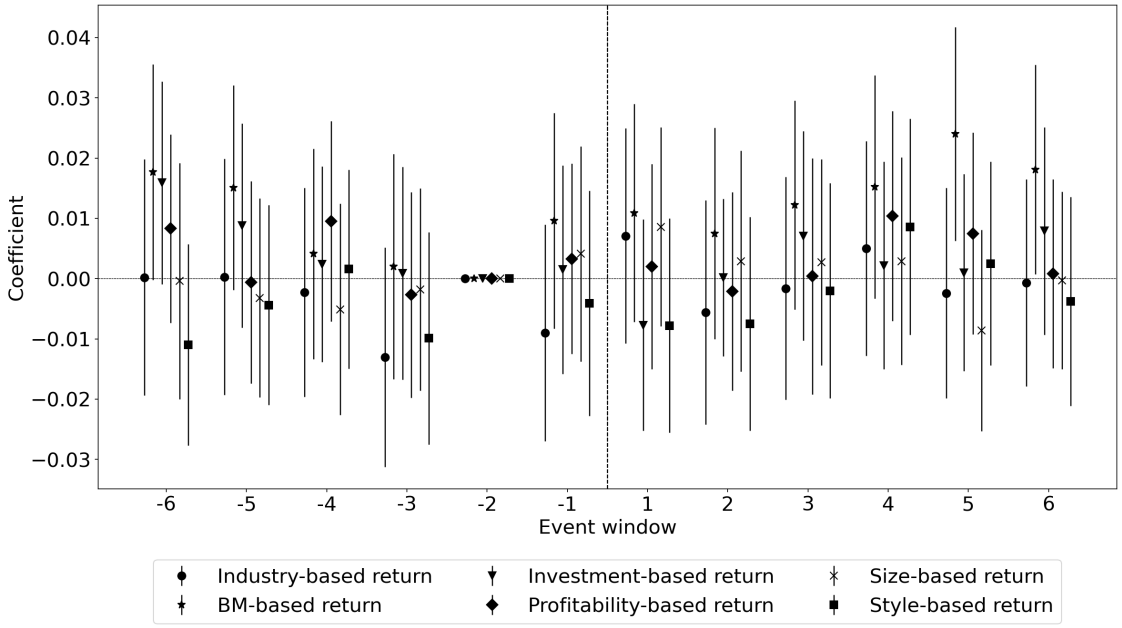


Figure 6: 2-for-1 and Moderate-Factor Splits: Alternative Characteristics
This figure presents the difference between 2-for-1 and 3-for-2 splits under six specifications: industry-based return (Moskowitz and Grinblatt, 1999), style-based return (Wahal and Yavuz, 2013), and four characteristic-based returns (size, book-to-market ratio, profitability, and investment). For each characteristic, we define peer stocks analogously as when defining price-based peers, then we calculate value-weighted average monthly return of the focal stocks' characteristic-based peer stocks. For each specification, the figure shows coefficients (γ_k) from regression (5):

$$\begin{aligned}\text{PredictPower}_{i,t+1} = &\sum_{k=-6}^{6} \beta_k \mathbb{I}\{EventMonth_{i,t} = k\} \\ &+ \sum_{k=-6}^{6} \gamma_k \mathbb{I}\{EventMonth_{i,t} = k\} \times \mathbb{I}\{TwoForOne_{i,t}\} \\ &+ \mathbb{I}\{TwoForOne_{i,t}\} + \alpha + \tau_t + \eta_i + \varepsilon_{i,t+1},\end{aligned}$$

where $k = -1$ serves as the estimation benchmark. The dependent variable $\text{PredictPower}_{i,t+1}$ is an indicator variable that equals one if both stock i's current monthly return $(t+1)$ and its lagged predictor (t) fall within either the top or bottom quintile, and zero otherwise. Dummy variable $\mathbb{I}\{EventMonth_{i,t} = k\}$ equals to one if month t is k months before or after a split event for stock i, with $k = 1$ denoting the split event month. $\mathbb{I}\{TwoForOne_{i,t}\}$ is an indicator variable that takes a value of one if the split event for stock i in month t is a 2-for-1 split, and zero if it is a 3-for-2 split or 5-for-4 split. The regression includes year-month fixed effects (τ_t) and firm fixed effects (η_i). Standard errors are clustered by stock and month. The sample period is from 1964 to 2024.

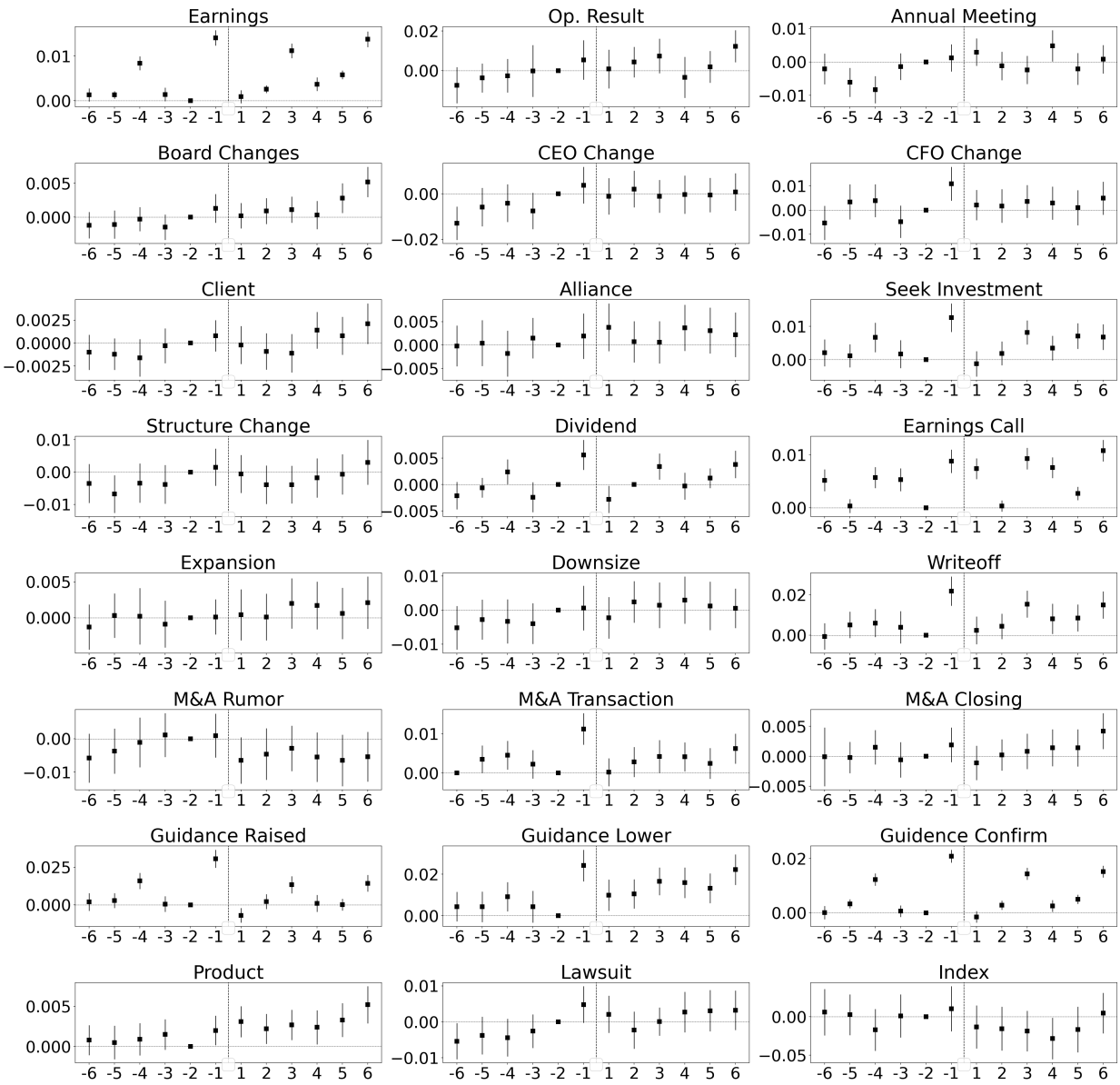


Figure 7: Placebo Events

This figure shows changes in the price-based return predictability around corporate events by presenting the regression coefficients (β_k) from equation (4):

$$\text{PredictPower}_{i,t+1} = \alpha + \sum_{k=-6}^{6} \beta_k \mathbb{I}\{EventMonth_{i,t} = k\} + \tau_t + \eta_i + \varepsilon_{i,t+1},$$

where $k = -1$ serves as the estimation benchmark. The dependent variable $\text{PredictPower}_{i,t+1}$ is an indicator variable that equals one if both stock i's current monthly return $(t+1)$ and its lagged RPSIM (t) fall within either the top or bottom quintile, and zero otherwise. Dummy variable $\mathbb{I}\{EventMonth_{i,t} = k\}$ equals to one if month t is k months before or after a corporate event for stock i, with $k = 1$ denoting the event month. The regression includes year-month fixed effects (τ_t) and firm fixed effects (η_i). Standard errors are clustered by stock and month.

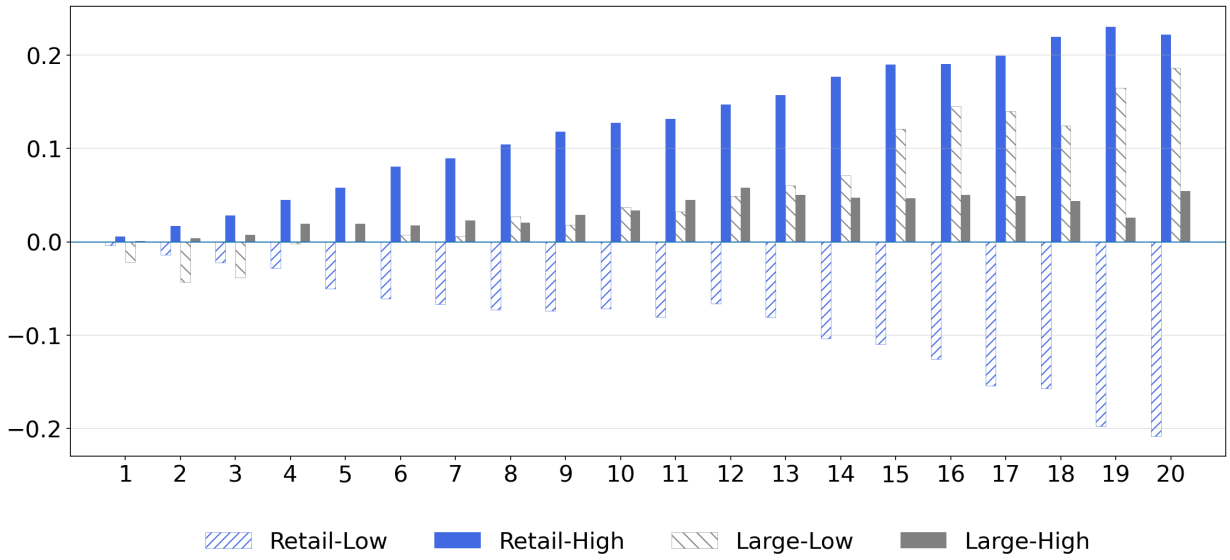


Figure 8: Subsequent Order Imbalance

This figure plots the detrended cumulative order imbalance in the subsequent trading month. At the end of each month t, we form decile portfolios based on RPSIM. For each stock, we measure the change in the retail (large) order imbalance by taking the difference between the daily imbalance in month $t+1$ and the average daily imbalance of the six-month window from month $t-7$ to month $t-2$. Then we calculate the cumulative retail order imbalance. To better demonstrate the results, we substract the average order imbalance from the order imbalance of decile groups in each trading day to detrend. For example, Retail-Low represents the average cumulative retail order imbalance for the bottom RPSIM decile portfolio. The sample period is from 1983 to 2000.

Table 1: Summary Statistics of RPSIM Portfolios

This table reports the time-series average of firm-level characteristics. Each month, we sort stocks into deciles based on RPSIM and compute the mean values of the characteristics listed in the top row of the table across all stocks in the decile for each portfolio. For each stock, we define PSIM as the value-weighted average monthly return of the 100 nearest price-comparable stocks, where the distance between stocks is calculated using nominal share prices. RPSIM is the residual of cross-sectional regression of PSIM on the focal stock's contemporaneous monthly return. Reported characteristics include the sorting variable RPSIM, RPSIM in the subsequent month (RPSIM_t), market beta (Beta), share price (Prc), the natural logarithm of market capitalization (Size), book-to-market ratio (BM), operating profit (Profit), asset growth (Invest), past 1-month return (R1m), past 12-month return skipping the most recent month (Mom), idiosyncratic volatility (IVol) of Ang et al. (2006), and the Amihud (2002) illiquidity measure (Illiq). The last row reports time-series averages of cross-sectional correlations between RPSIM and various characteristics. All variables are winsorized each month at the 1% and 99% levels. The sample period is from June 1963 to December 2024.

Decile	RPSIM	RPSIM_{t+1}	Beta	Prc	Size	BM	Profit	Invest	R1m	Mom	IVol	Illiq
High	5.772	1.050	1.143	37.303	12.872	0.707	0.367	0.177	0.640	17.963	0.027	2.442
9	3.641	1.067	1.135	33.858	12.919	0.704	0.366	0.175	1.355	17.430	0.026	1.856
8	2.510	0.865	1.140	30.212	12.795	0.717	0.364	0.176	1.300	16.295	0.026	1.774
7	1.577	0.660	1.150	27.297	12.627	0.730	0.364	0.177	1.303	15.378	0.026	1.887
6	0.690	0.387	1.162	25.128	12.455	0.748	0.363	0.178	1.173	14.297	0.028	2.142
5	-0.215	0.114	1.174	22.603	12.242	0.765	0.361	0.181	0.939	12.558	0.029	2.679
4	-1.184	-0.221	1.185	20.317	12.002	0.791	0.358	0.177	0.781	11.165	0.031	3.423
3	-2.305	-0.646	1.201	17.775	11.713	0.821	0.352	0.174	0.550	8.650	0.034	4.697
2	-3.781	-1.184	1.210	14.654	11.335	0.863	0.345	0.164	0.401	4.900	0.038	6.705
Low	-6.691	-1.938	1.217	9.890	10.763	0.922	0.328	0.148	1.467	-1.973	0.045	10.973
Corr. RPSIM	1.000	0.232	-0.039	0.241	0.314	-0.110	0.039	0.010	0.009	0.109	-0.276	-0.126

Table 2: Portfolio Returns by Sorting on RPSIM

This table reports the monthly returns and alphas of portfolios sorted on RPSIM. For each stock at the end of each month t, we calculate the Euclidean distance to all other stocks based on nominal share price. PSIM is the value-weighted average excess (to the risk-free rate) return of the nearest 100 price-comparable in the past month. RPSIM is the residual from cross-sectional regressions of PSIM on the focal stock's own monthly return. Panel A reports the average excess returns and alphas of the top and the bottom decile portfolios and the long-short strategy. In Panel B, RPSIM is calculated based on the value-weighted average excess return of the past year or the past week; portfolios are value-weighted. Panel C considers alternative specifications of price-comparable stocks, including price-based peers within the same Fam-MacBeth industry classification and price-based peers from different industries, price-based peers excluding shared analyst coverage (Ali and Hirshleifer, 2020), and randomly matched peer stocks; returns are value-weighted throughout Panel C. Panel D reports portfolio performance based on the original PSIM variable. To calculate alpha, we consider the capital asset pricing model (CAPM), the Fama and French (1996) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the momentum-augmented Fama and French (2015) model (FF6). Returns are reported in %. The sample period is from July 1963 to December 2024. Newey-West three-lag adjusted t-statistics are reported in parentheses.

Panel A. Baseline: portfolios based on RPSIM

	Value Weighted			Equal Weighted		
	High	Low	High-Low	High	Low	High-Low
Excess Return	1.08	-0.05	1.13	1.20	0.25	0.95
	(5.18)	(-0.16)	(5.52)	(4.76)	(0.80)	(6.46)
CAPM Alpha	0.46	-0.81	1.27	0.53	-0.46	0.99
	(4.69)	(-4.74)	(6.58)	(3.94)	(-2.30)	(6.85)
FF3 Alpha	0.48	-0.90	1.38	0.47	-0.55	1.02
	(5.53)	(-6.22)	(7.56)	(5.78)	(-6.98)	(10.03)
FF5 Alpha	0.55	-0.75	1.31	0.53	-0.40	0.93
	(5.93)	(-4.53)	(5.91)	(6.07)	(-2.91)	(6.00)
FF6 Alpha	0.58	-0.52	1.10	0.53	-0.19	0.72
	(5.86)	(-3.09)	(4.84)	(5.62)	(-1.27)	(4.20)

Panel B. Alternative reference periods when constructing RPSIM

	Yearly return-based RPSIM			Weekly return-based RPSIM		
	High	Low	High-Low	High	Low	High-Low
Excess Return	0.69	0.19	0.50	1.11	-0.05	1.16
	(3.47)	(0.57)	(1.97)	(4.48)	(-0.18)	(6.19)
CAPM Alpha	0.09	-0.59	0.68	0.42	-0.77	1.19
	(0.85)	(-2.94)	(2.88)	(3.38)	(-5.16)	(6.51)
FF3 Alpha	0.12	-0.70	0.82	0.38	-0.84	1.22
	(1.25)	(-4.30)	(3.86)	(3.24)	(-6.60)	(6.80)
FF5 Alpha	0.12	-0.50	0.62	0.47	-0.77	1.24
	(1.15)	(-2.94)	(2.73)	(3.64)	(-6.30)	(6.76)
FF6 Alpha	-0.02	-0.16	0.14	0.65	-0.64	1.29
	(-0.21)	(-1.01)	(0.67)	(4.73)	(-5.14)	(6.11)

Panel C. Alternative specifications of price-comparable stocks						
	Same Industry			Different Industry		
	High	Low	High-Low	High	Low	High-Low
Excess Return	0.93	-0.24	1.18	0.92	0.16	0.76
	(4.98)	(-1.07)	(7.00)	(4.51)	(0.58)	(3.74)
CAPM Alpha	0.37	-0.92	1.29	0.31	-0.59	0.89
	(3.90)	(-7.43)	(7.67)	(3.14)	(-3.53)	(4.56)
FF3 Alpha	0.36	-0.93	1.28	0.29	-0.68	0.97
	(3.64)	(-7.92)	(7.66)	(3.20)	(-4.69)	(5.34)
FF5 Alpha	0.39	-0.87	1.27	0.34	-0.53	0.87
	(3.43)	(-7.28)	(6.73)	(3.40)	(-3.12)	(4.43)
FF6 Alpha	0.30	-0.81	1.11	0.40	-0.33	0.73
	(2.93)	(-6.56)	(6.23)	(4.11)	(-1.87)	(3.43)
	Exclude common analyst			Randomly matched peers		
	High	Low	High-Low	High	Low	High-Low
Excess Return	0.97	-0.04	1.01	0.51	0.64	-0.12
	(4.72)	(-0.14)	(5.02)	(2.90)	(3.64)	(-1.47)
CAPM Alpha	0.36	-0.79	1.16	-0.07	0.04	-0.10
	(3.66)	(-4.77)	(6.05)	(-1.06)	(0.64)	(-1.24)
FF3 Alpha	0.36	-0.89	1.25	-0.05	0.06	-0.11
	(4.03)	(-6.30)	(7.11)	(-0.80)	(1.04)	(-1.31)
FF5 Alpha	0.40	-0.76	1.16	-0.06	0.05	-0.11
	(4.15)	(-4.60)	(5.99)	(-0.96)	(0.82)	(-1.27)
FF6 Alpha	0.44	-0.53	0.97	-0.05	0.06	-0.11
	(4.63)	(-3.12)	(4.58)	(-0.74)	(1.01)	(-1.21)
Panel D. Original PSIM without orthogonalization						
	Value Weighted			Equal Weighted		
	High	Low	High-Low	High	Low	High-Low
Excess Return	1.05	0.13	0.92	1.11	0.45	0.66
	(5.08)	(0.42)	(4.22)	(4.51)	(1.41)	(4.24)
CAPM Alpha	0.44	-0.66	1.10	0.46	-0.27	0.73
	(4.31)	(-3.67)	(5.30)	(3.51)	(-1.37)	(4.84)
FF3 Alpha	0.46	-0.76	1.21	0.40	-0.37	0.78
	(4.81)	(-4.90)	(6.13)	(4.94)	(-2.79)	(5.38)
FF5 Alpha	0.55	-0.59	1.13	0.46	-0.21	0.67
	(5.04)	(-3.23)	(4.52)	(5.31)	(-1.45)	(3.97)
FF6 Alpha	0.56	-0.33	0.89	0.45	0.02	0.43
	(5.24)	(-1.67)	(3.46)	(4.76)	(0.11)	(2.27)

Table 3: Double-Sorted Portfolio Returns

This table reports double-sorted portfolio returns based on RPSIM and confounding factors. Size is the natural logarithm of market capitalization. BM is the book-to-market ratio. Profit is the operating profitability (Fama and French, 2015). Invest is asset growth (Fama and French, 2015). R1m is the past 1-month return of the focal stock. Mom is the cumulative return from month $t-12$ to month $t-2$. Ltrev is the cumulative return from month $t-60$ to month $t-13$. IVol is idiosyncratic volatility (Ang et al., 2006). MAX is the maximum daily return in month $t-1$ (Bali et al., 2011). Skew is the skewness of monthly returns over the previous five years. Coskew is coskewness (Harvey and Siddique, 2000) using five years of monthly returns. IndRet represents the one-month industry momentum (Moskowitz and Grinblatt, 1999). Each month, stocks are sorted into quintiles based on a confounding factor; within each quintile, stocks are then further sorted into deciles based on RPSIM. The returns of the decile portfolios over the next month are averaged across the five control variable quintiles. We report the five-factor model (Fama and French, 2015) alphas, on a value-weighted basis, of the decile RPSIM portfolios and of the high-RPSIM minus low-RPSIM long-short portfolio. The sample period runs from July 1963 to December 2024. Newey-West three-lag adjusted t-statistics are reported in parentheses.

	Size	BM	Profit	Invest	R1m	Mom	Ltrev	IVol	MAX	Skew	Coskew	IndRet
High	0.24	0.32	0.46	0.52	0.51	0.34	0.39	0.32	0.44	0.45	0.47	0.56
	(4.56)	(4.31)	(5.34)	(4.80)	(5.29)	(4.00)	(4.99)	(3.08)	(4.73)	(5.34)	(4.79)	(6.23)
9	0.15	0.14	0.19	0.13	0.16	0.11	0.14	0.09	0.18	0.12	0.23	0.18
	(3.25)	(2.20)	(2.56)	(2.08)	(2.32)	(1.52)	(2.05)	(0.94)	(2.23)	(1.66)	(3.22)	(2.44)
8	0.12	0.04	0.09	0.05	0.04	-0.02	0.06	0.04	-0.00	0.08	0.08	0.14
	(2.45)	(0.70)	(1.49)	(0.75)	(0.67)	(-0.24)	(1.10)	(0.52)	(-0.03)	(1.39)	(1.56)	(2.27)
7	0.00	0.04	-0.05	0.06	0.16	-0.16	0.06	-0.09	-0.05	-0.02	0.10	-0.02
	(0.03)	(0.69)	(-0.80)	(0.88)	(2.22)	(-2.51)	(0.94)	(-1.02)	(-0.68)	(-0.40)	(1.48)	(-0.30)
6	0.09	-0.02	-0.08	-0.11	-0.16	-0.17	-0.04	-0.04	-0.14	0.00	-0.14	-0.15
	(1.87)	(-0.39)	(-1.15)	(-1.87)	(-2.50)	(-2.36)	(-0.72)	(-0.58)	(-1.86)	(0.03)	(-2.02)	(-2.46)
5	-0.05	-0.22	-0.13	-0.17	-0.09	-0.23	-0.17	-0.18	-0.24	-0.18	-0.12	-0.21
	(-0.87)	(-3.40)	(-2.00)	(-2.41)	(-1.22)	(-3.27)	(-2.75)	(-2.31)	(-3.25)	(-2.53)	(-1.82)	(-3.30)
4	-0.14	-0.27	-0.30	-0.15	-0.18	-0.29	-0.23	-0.24	-0.22	-0.28	-0.24	-0.30
	(-2.13)	(-4.00)	(-3.74)	(-2.19)	(-1.93)	(-4.04)	(-3.58)	(-3.01)	(-2.91)	(-3.89)	(-3.20)	(-4.11)
3	-0.08	-0.32	-0.47	-0.48	-0.33	-0.36	-0.36	-0.29	-0.41	-0.37	-0.44	-0.42
	(-1.22)	(-3.77)	(-4.88)	(-4.97)	(-3.43)	(-4.06)	(-5.18)	(-3.38)	(-4.89)	(-4.64)	(-4.39)	(-4.24)
2	-0.16	-0.40	-0.41	-0.40	-0.41	-0.46	-0.40	-0.36	-0.39	-0.44	-0.33	-0.44
	(-2.08)	(-3.72)	(-3.28)	(-3.48)	(-3.81)	(-5.11)	(-4.44)	(-3.42)	(-3.93)	(-4.80)	(-3.43)	(-4.16)
Low	-0.31	-0.56	-0.56	-0.49	-0.57	-0.41	-0.42	-0.37	-0.37	-0.43	-0.58	-0.71
	(-3.58)	(-3.67)	(-3.47)	(-2.95)	(-3.46)	(-3.14)	(-2.66)	(-3.68)	(-2.86)	(-3.11)	(-3.53)	(-4.71)
High-Low	0.55	0.88	1.02	1.00	1.08	0.75	0.80	0.69	0.81	0.88	1.05	1.27
	(5.74)	(5.02)	(4.93)	(4.37)	(5.09)	(4.63)	(4.35)	(4.79)	(4.86)	(5.02)	(4.82)	(6.20)

Table 4: Fama-MacBeth Regressions

This table reports time-series averages of coefficients from monthly cross-sectional regressions, where the dependent variable one month-ahead stock return (in %). The main independent variable of interest, RPSIM, is the residual average return of the focal stock's price-based peers. In column (1), only RPSIM is included. In column (2), we control for one-month style and industry returns (Style 1M and Industry 1M). In column (3), we control for six-month style and industry returns (Style 6M and Industry 6M). In column (4), we control for 12-month style and industry returns (Style 12M and Industry 12M). Column (5) include style and industry returns across all horizons. Other firm-level characteristics include size, book-to-market ratio, operating profitability, asset growth, idiosyncratic volatility (Ang et al., 2006), cumulative return from month $t-12$ to month $t-2$, illiquidity (Amihud, 2002), and the focal stock's share price. Independent variables are standardized each month to have zero mean and unit variance. The sample period is from June 1963 to December 2024. The last three rows in the table report the average R^2 for cross-sectional regressions, the average number of observations each month, and the number of months. Newey-West three-lag adjusted t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)
RPSIM	0.164	0.174	0.177	0.176	0.171
	(5.83)	(5.91)	(6.01)	(5.98)	(5.95)
StyleRet 1m		0.024			-0.051
		(0.87)			(-1.89)
IndRet 1m		0.249			0.188
		(7.00)			(5.70)
StyleRet 6m			0.056		-0.011
			(1.73)		(-0.22)
IndRet 6m			0.234		0.105
			(6.14)		(1.83)
StyleRet 12m				0.081	0.151
				(2.28)	(2.92)
IndRet 12m				0.233	0.097
				(6.09)	(1.94)
Controls	Yes	Yes	Yes	Yes	Yes
Intercept	0.780	0.783	0.781	0.783	0.782
	(3.27)	(3.29)	(3.28)	(3.29)	(3.29)
Avg. R^2(%)	5.046	5.579	5.588	5.591	6.180
Avg. Obs.	2535	2431	2431	2431	2431
# Months	738	738	738	738	738

Table 5: Comovement and Predictability In the Green and Hwang (2009) Setting

This table reports average slope coefficients before and after stock splits, along with the corresponding changes. For each split event, Panel A estimates comovement regressions as follows:

$$R_{i,d} = \alpha_i + \beta_{LowPrc,i} R_{LowPrc,i,d} + \beta_{HighPrc,i} R_{HighPrc,i,d} + \varepsilon_{i,d},$$

where $R_{i,d}$ is the return of stock i on day d and $R_{LowPrc,i,d}$ ($R_{HighPrc,i,d}$) is stock i's respective low (high) price index on day d. Panel B estimates predictive regressions as follows:

$$R_{i,d+1} = \alpha_i + \beta_{LowPrc,i} R_{LowPrc,i,d} + \beta_{HighPrc,i} R_{HighPrc,i,d} + \beta_{Focal,i} R_{i,d} + \varepsilon_{i,d+1},$$

where the focal stock's lagged return $R_{i,d}$ is included to control for the short-term reversal effect. In each regression, the low-price index comprises stocks with prices in $[\frac{1}{4}P_{pre}, \frac{3}{4}P_{pre}]$, and the high-price index consists of stocks with prices in $[\frac{3}{4}P_{pre}, \frac{5}{4}P_{pre}]$, where P_{pre} is the pre-split price measured one day before the split. The sample includes 2-for-1 split events from 1964 to 2024. The table reports the average $\beta_{LowPrc,i}$, the average $\beta_{HighPrc,i}$, and the average difference $\beta_{HighPrc,i} - \beta_{LowPrc,i}$ in the pre- and post-split periods, as well as the change between the two periods. Standard errors are clustered by month, and t-statistics are reported in parentheses.

Panel A. Contemporaneous comovement			
	Before splits	After splits	Difference: After−Before
LowPrc	0.812	1.018	0.206
	(42.54)	(50.29)	(9.01)
HighPrc	0.213	0.127	-0.085
	(10.59)	(6.60)	(-4.39)
Difference:	-0.599	-0.891	-0.291
HighPrc−LowPrc	(-16.34)	(-24.25)	(-7.11)
Panel B. Predictive regressions			
	Before splits	After splits	Difference: After−Before
LowPrc	0.063	0.144	0.081
	(3.90)	(9.24)	(3.68)
HighPrc	0.168	0.117	-0.051
	(10.87)	(7.33)	(-2.35)
Difference:	0.105	-0.027	-0.132
HighPrc−LowPrc	(3.43)	(-0.89)	(-3.10)

Table 6: Price-Comparable Stock Returns and Retail Investor Attention

This table reports panel regressions of retail investor attention on lagged RPSIM. For each stock, the level of retail investor attention is measured by abnormal Google search volume (Da et al., 2011). In Panel A, the main independent variable of interest is the absolute value of RPSIM:

$$\text{Retail Attention}_{i,t+1} = \alpha + \beta|RPSIM_{i,t}| + Controls + \varepsilon_{i,t+1}.$$

In Panel B, we examine the positive and the negative parts of RPSIM separately:

$$\text{Retail Attention}_{i,t+1} = \alpha + \beta_1 RPSIM^{+}_{i,t} + \beta_2 RPSIM^{-}_{i,t} + Controls + \varepsilon_{i,t+1},$$

where $RPSIM^{+}_{i,t} = \max\{RPSIM^{+}_{i,t}, 0\}$ and $RPSIM^{-}_{i,t} = \min\{RPSIM^{+}_{i,t}, 0\}$. Control variables, defined identically as in Table 3, include firm size, book-to-market ratio, operating profitability, investment, idiosyncratic volatility, momentum, illiquidity, MAX, Skew, Ltrev, Coskew, and share price. We also control for the focal stock's absolute monthly return. In columns (5) and (6), we further control for lagged retail investor attention. We also include industry fixed effect and the month fixed effect. Standard errors are clustered by firm and month, and t-statistics are reported in parentheses. The sample period is from April 2004 to December 2020.

	(1)	(2)	(3)	(4)	(5)
Panel A. Absolute RPSIM					
|RPSIM|	0.187	0.202	0.198	0.189	0.185
	(3.81)	(3.61)	(3.55)	(3.37)	(3.31)
Retail Attention				0.019	0.018
				(3.84)	(3.73)
Controls	No	Yes	Yes	Yes	Yes
Industry FE	Yes	No	Yes	No	Yes
Year-month FE	Yes	Yes	Yes	Yes	Yes
Observations	485297	459465	444690	456561	441802
Adjusted R^2	0.00	0.01	0.00	0.01	0.01
Panel B. Positive and negative parts of RPSIM					
RPSIM^{+}	0.186	0.206	0.214	0.206	0.215
	(2.50)	(2.64)	(2.72)	(2.58)	(2.66)
RPSIM^{-}	-0.187	-0.198	-0.185	-0.175	-0.162
	(-3.52)	(-3.24)	(-3.02)	(-3.04)	(-2.81)
Retail Attention				0.019	0.018
				(3.84)	(3.73)
Controls	No	Yes	Yes	Yes	Yes
Industry FE	Yes	No	Yes	No	Yes
Year-month FE	Yes	Yes	Yes	Yes	Yes
Observations	485297	459465	444690	456561	441802
Adjusted R^2	0.00	0.01	0.00	0.01	0.01

Table 7: Subsample Analysis

This table reports the monthly alphas on value-weighted quintile portfolios sorted on RPSIM. In Panel A, we divide the whole sample into two subperiods: 1963 to 1989, and 1990 to 2024. In Panel B, we split the sample based on the median market size breakpoint of NYSE stocks. In Panel C, subsamples are defined based on the median idiosyncratic volatility (IVol). In Panel D, we split the sample into two groups based on institutional ownership (IO). We use the size-adjusted institutional ownership following the method of Nagel (2005). Newey-West three-lag adjusted t-statistics are reported in parentheses. The sample period in Panel B, Panel C, and Panel D is from July 1963 to December 2024.

		Low	2	3	4	High	High-Low
Panel A. Earlier and recent sample periods							
1963-1989	Excess Return	-0.30	0.19	0.27	0.56	0.86	1.16
		(-0.81)	(0.59)	(0.97)	(1.97)	(2.98)	(5.07)
	FF3 Aplha	-0.95	-0.41	-0.24	0.07	0.40	1.35
		(-6.70)	(-4.09)	(-3.26)	(1.01)	(4.20)	(7.19)
	FF6 Aplha	-0.64	-0.14	-0.10	0.10	0.35	0.99
		(-3.72)	(-1.23)	(-1.35)	(1.38)	(3.43)	(4.27)
1990-2024	Excess Return	0.33	0.44	0.60	0.80	0.97	0.64
		(0.93)	(1.65)	(2.41)	(3.41)	(3.72)	(2.66)
	FF3 Aplha	-0.63	-0.37	-0.17	0.06	0.22	0.85
		(-3.62)	(-3.51)	(-2.17)	(0.81)	(2.27)	(3.88)
	FF6 Aplha	-0.32	-0.30	-0.09	0.06	0.30	0.62
		(-1.67)	(-2.93)	(-1.20)	(0.77)	(2.74)	(2.33)
Panel B. Small and large subsamples							
Size<NYSE p50	Excess Return	0.18	0.46	0.68	0.82	1.02	0.85
		(0.58)	(1.69)	(2.62)	(3.25)	(3.99)	(6.50)
	FF3 Aplha	-0.72	-0.41	-0.17	0.00	0.21	0.93
		(-6.79)	(-6.20)	(-3.47)	(0.07)	(3.66)	(7.76)
	FF6 Aplha	-0.35	-0.18	-0.03	0.12	0.32	0.68
		(-2.93)	(-2.16)	(-0.52)	(2.78)	(4.83)	(4.60)
Size>NYSE p50	Excess Return	0.39	0.53	0.55	0.63	0.71	0.32
		(2.15)	(3.14)	(3.31)	(3.94)	(4.06)	(3.26)
	FF3 Aplha	-0.22	-0.04	0.00	0.12	0.17	0.38
		(-3.39)	(-0.83)	(0.04)	(2.42)	(3.17)	(4.06)
	FF6 Aplha	-0.14	-0.06	-0.02	0.11	0.20	0.34
		(-2.04)	(-1.08)	(-0.35)	(2.08)	(3.63)	(3.34)

		Low	2	3	4	High	High-Low
Panel C. High and low idiosyncratic volatility subsamples							
High IVol	Excess Return	0.17	0.37	0.53	0.71	0.80	0.64
		(0.48)	(1.14)	(1.64)	(2.24)	(2.46)	(3.30)
	FF3 Aplha	-0.78	-0.55	-0.40	-0.15	-0.01	0.77
		(-5.20)	(-3.84)	(-2.71)	(-1.15)	(-0.04)	(4.32)
	FF6 Aplha	-0.28	-0.10	-0.04	0.22	0.36	0.64
		(-1.72)	(-0.67)	(-0.32)	(1.71)	(2.34)	(2.70)
Low IVol	Excess Return	0.36	0.56	0.53	0.66	0.81	0.46
		(2.06)	(3.30)	(3.26)	(4.19)	(4.98)	(4.26)
	FF3 Aplha	-0.26	-0.04	-0.04	0.11	0.26	0.52
		(-3.39)	(-0.64)	(-0.77)	(2.07)	(4.96)	(5.03)
	FF6 Aplha	-0.23	-0.05	-0.06	0.06	0.22	0.46
		(-2.88)	(-0.82)	(-1.13)	(1.07)	(4.08)	(4.02)
Panel D. High and low institutional ownership subsamples							
High IO	Excess Return	0.18	0.46	0.59	0.77	0.90	0.72
		(0.64)	(2.04)	(2.79)	(3.82)	(4.22)	(4.36)
	FF3 Aplha	-0.65	-0.28	-0.15	0.05	0.20	0.85
		(-4.83)	(-3.14)	(-1.88)	(0.85)	(2.86)	(5.20)
	FF6 Aplha	-0.54	-0.28	-0.15	0.03	0.21	0.75
		(-4.09)	(-3.24)	(-2.08)	(0.36)	(2.96)	(4.40)
Low IO	Excess Return	0.06	0.37	0.63	0.90	1.04	0.98
		(0.18)	(1.25)	(2.36)	(3.39)	(3.48)	(4.20)
	FF3 Aplha	-0.78	-0.46	-0.14	0.21	0.32	1.10
		(-4.37)	(-3.14)	(-1.25)	(1.74)	(1.89)	(4.86)
	FF6 Aplha	-0.30	-0.04	0.15	0.43	0.56	0.86
		(-1.63)	(-0.21)	(1.33)	(3.73)	(3.41)	(3.19)

Table 8: International Evidence on Price-Based Return Predictability

This table reports the average returns of portfolios formed on RPSIM in 38 markets. Considering the limited availability of stocks in certain countries, we calculate the RPSIM by uniformly utilizing the residual average returns of 30 nearest price-comparable stocks. We divide sample into two groups based on aggregate turnover, which is the average annual turnover across all the years for which we have data for the market. For turnover higher than the cross-sectional median level, we use the past one-week residual average return of peer stocks as our predictive variable RPSIM, whereas for turnover below the median, we still use the monthly returns. Also reported are returns on a strategy that equally combines the strategies available within a given week or month (Comb).

	High turnover				Low turnover		
Market	Low	High	High-Low	Market	Low	High	High-Low
AUS	-0.021	0.284	**0.305**	ARG	1.426	0.785	-0.642
	(-0.22)	(3.39)	(5.30)		(2.20)	(1.20)	(-1.16)
BRA	0.325	0.292	-0.032	AUT	0.467	0.870	0.402
	(1.86)	(1.96)	(-0.20)		(1.03)	(1.77)	(0.94)
CHE	0.187	0.199	0.012	BEL	0.059	0.544	0.485
	(2.62)	(2.81)	(0.19)		(0.17)	(1.62)	(1.50)
CHN	-0.035	0.237	**0.271**	CAN	0.749	1.464	**0.716**
	(-0.33)	(2.34)	(4.06)		(1.55)	(3.60)	(1.85)
DEU	0.060	0.109	0.049	CHL	0.645	0.778	0.133
	(0.59)	(1.35)	(0.48)		(1.74)	(2.07)	(0.63)
GBR	0.117	0.059	-0.058	DNK	0.575	0.599	0.024
	(1.54)	(0.84)	(-0.95)		(1.60)	(1.62)	(0.07)
HKG	-0.013	0.182	**0.195**	ESP	0.180	0.661	0.481
	(-0.14)	(1.87)	(2.61)		(0.52)	(2.01)	(1.58)
IND	0.156	0.441	**0.284**	FIN	0.792	1.110	0.318
	(1.47)	(4.23)	(3.82)		(1.90)	(2.61)	(0.90)
ISR	0.226	0.243	0.017	FRA	0.371	0.519	0.148
	(2.50)	(2.85)	(0.29)		(1.09)	(1.58)	(0.61)
ITA	0.082	0.179	0.097	GRC	0.381	0.490	0.108
	(0.99)	(2.29)	(1.60)		(0.63)	(0.90)	(0.24)
JPN	0.059	0.154	**0.096**	IDN	0.944	0.829	-0.115
	(0.79)	(2.13)	(2.71)		(1.31)	(1.26)	(-0.29)
KOR	0.018	0.086	0.067	MEX	0.448	1.137	**0.689**
	(0.14)	(0.68)	(0.86)		(1.10)	(2.72)	(2.12)
NLD	0.037	0.226	**0.190**	MYS	0.359	0.449	0.090
	(0.42)	(2.78)	(2.33)		(0.65)	(0.98)	(0.32)
NOR	0.022	0.281	**0.259**	NZL	0.213	0.699	0.486
	(0.21)	(2.69)	(2.96)		(0.59)	(2.15)	(1.94)
PAK	0.059	0.363	**0.304**	PHL	0.715	0.570	-0.145
	(0.55)	(3.39)	(3.41)		(1.21)	(1.24)	(-0.42)
RUS	-0.007	0.293	**0.300**	POL	0.440	1.263	**0.823**
	(-0.05)	(1.64)	(1.73)		(0.88)	(2.54)	(2.44)
SWE	0.090	0.325	**0.235**	PRT	1.239	0.377	-0.862
	(0.91)	(3.48)	(3.23)		(1.17)	(0.40)	(-0.93)
THA	0.113	0.240	0.127	SGP	0.448	0.722	0.273
	(0.98)	(2.28)	(1.77)		(0.95)	(1.60)	(0.83)
TUR	0.173	0.285	0.112	ZAF	0.570	0.875	0.305
	(1.11)	(1.82)	(1.35)		(1.26)	(1.93)	(0.99)
Comb	0.084	0.235	**0.151**	Comb	0.552	0.792	**0.241**
	(1.43)	(4.03)	(7.71)		(1.84)	(2.77)	(2.95)

Appendix

A Retail Order Imbalance

The measure of retail order imbalance (RIMB) is inferred from the transaction data from ISSM and TAQ. We employ the filter from Hvidkjaer (2006) and only include NYSE and AMEX common stocks from 1983 to 2000.[16] Trades and quotes with irregular terms are excluded, and likely erroneous prices are excluded by a simple price-based error filter.

To begin with, we follow the algorithm from Lee (1992) to classify all trades by size. Specifically, at the end of month $t-1$, all stocks are grouped based on their firm size. Then, all the trades in month t are classified as small, medium, or large trade using the following dollar value breakpoints for each size-quintile group:

Firm-size quintile	Small	2	3	4	Large
Small trade cutoff, in $	3400	4800	7300	10300	16400
Large trade cutoff, in $	6800	9600	14600	20600	32800

All trades are further classified as buy-initiated or sell-initiated based on the Lee and Ready (1991) algorithm. This standard procedure uses trade price relative to the current bid-ask quotes to determine trade direction. A trade is classified as a buy-initiated (sell-initiated) if it is executed at a price above (below) the quote midpoint. If it is executed right at the quote midpoint, then we classify the trade as buy-initiated (sell-initiated) if the trade price is above (below) the last executed trade price.

According to the previous two steps, all trades are classified by size and direction. Retail order imbalance (RIMB) is then computed as the relative difference between buy-initiated and sell-initiated trade volume in a small trade group:

$$RIMB = \frac{BUYVOL - SELLVOL}{BUYVOL + SELLVOL}$$

For each stock, we measure the change in the retail (large) order imbalance by taking the difference between the daily imbalance and the average daily imbalance over the six-month window from month $t-7$ to month $t-2$. Then, we calculate the cumulative retail order imbalance, and to

[16] Following Barber et al. (2008), our retail order imbalance sample ends in 2000 since the widespread introduction of decimalization in 2001 impacts significantly the distribution of trade size.

better demonstrate the results, we substract the average order imbalance from the order imbalance of the corresponding portfolio in each trading day (or month) to detrend.

B Additional Empirical Results

Table A1: Portfolio Returns by Sorting on RPSIM: Alternative Factor Models

This table reports the monthly returns and alphas of portfolios sorted on RPSIM. For each stock at the end of each month t, we calculate the Euclidean distance to all other stocks based on nominal share price. PSIM is the value-weighted average excess (to the risk-free rate) return of the nearest 100 stocks in the past month. RPSIM is the residual from cross-sectional regressions of PSIM on the focal stock's own monthly return. The reports alphas of the top and the bottom decile portfolios and the long-short strategy. We consider the Carhart (1997) four-factor model (Carhart), the Hou et al. (2015) Q-factor model (Q4), the Stambaugh and Yuan (2017) mispricing factor model (SY4), the Daniel et al. (2020) behavioral factor model (DHS3), and the Fama and French (2015) model augmented by the momentum factor and the short-term reversal factor (FF7). Returns are reported in %. The sample period is from July 1963 to December 2024. Newey-West three-lag adjusted t-statistics are reported in parentheses.

	Value Weighted			Equal Weighted		
	High	Low	High-Low	High	Low	High-Low
Carhart Alpha	0.51	-0.62	1.13	0.47	-0.30	0.77
	(5.44)	(-3.69)	(5.21)	(5.17)	(-2.02)	(4.77)
Q4 Alpha	0.63	-0.46	1.09	0.56	-0.06	0.61
	(5.95)	(-2.29)	(4.15)	(6.18)	(-0.34)	(3.27)
SY4 Alpha	0.40	-0.39	0.79	0.44	-0.13	0.58
	(3.66)	(-1.95)	(3.11)	(4.22)	(-0.70)	(2.95)
DHS3 Alpha	0.68	-0.12	0.80	1.02	0.44	0.58
	(5.77)	(-0.55)	(3.08)	(6.90)	(1.78)	(2.68)
FF7 Alpha	0.68	-0.68	1.36	0.54	-0.28	0.82
	(6.59)	(-4.13)	(6.01)	(5.28)	(-2.04)	(5.09)

Table A2: Portfolios by Sorting on RPSIM: Predicting Turnover

This table reports next-month average abnormal turnover for decile portfolios formed on RPSIM. For each stock at the end of each month t, we calculate the Euclidean distance to all other stocks based on nominal share price. PSIM is the value-weighted average excess return of the nearest 100 stocks in the past month. RPSIM is the residual from cross-sectional regressions of PSIM on the focal stock's own monthly return. Each month, we sort stocks into deciles by RPSIM and compute the value-weighted average abnormal turnover of each portfolio in the next month. Abnormal turnover is defined as monthly turnover minus the average turnover over the prior 3, 6, or 12 months. Turnovers are multiplied by 100. The sample period is from July 1963 to December 2024. Newey-West adjusted t-statistics are reported in parentheses.

	Abnormal Turnover (3m)	Abnormal Turnover (6m)	Abnormal Turnover (12m)
Low	-0.59	-1.90	-0.73
2	-0.49	-4.87	-0.41
3	-0.15	-2.17	-0.14
4	-0.19	-1.69	-0.16
5	-0.09	-1.48	-0.06
6	-0.05	-0.54	0.02
7	-0.02	-0.84	0.04
8	0.03	-0.64	0.10
9	0.10	-0.50	0.20
High	0.13	-0.69	0.32
High-Low	0.72	1.21	1.05
	(3.37)	(1.83)	(4.04)