

Margin Trading, Short Selling, and Return Predictability: Evidence from the Chinese Stock Market*

Guohao Tang^{a,b,c}, Hanrui Xu^{a,*}, Jing Yang^d, Fanyining Zhang^a

^a*College of Finance and Statistics, Hunan University, Changsha, China*

^b*Key Laboratory of High-Performance Distributed Ledger Technology and Digital Finance
(Ministry of Education), Changsha, 410006, China*

^c*Hunan Province China-Characteristic Financial Research
Center, Changsha, 410006, China*

^d*Xinjiang University of Science and Technology, Korla, China*

Abstract

Using Chinese A-share stocks eligible for margin trading and short selling from 2010 to 2025, this study constructs measures of margin-financing and short-selling activity along three dimensions: changes in outstanding balances, relative trading intensity, and sentiment momentum. Principal component analysis is then used to construct a composite factor capturing the relative strength of margin financing versus short selling. The factor significantly predicts lower future stock returns. A long-short portfolio that is long low-factor stocks and short high-factor stocks earns significant positive returns at both daily and weekly frequencies, and the predictive relation remains robust after controlling for conventional risk factors and firm characteristics. Predictive power declines as the holding horizon lengthens. Return predictability is stronger when investor sentiment is more intense or optimistic and among stocks with stronger lottery-like characteristics, consistent with short-term mispricing and subsequent price correction. The main findings remain robust after accounting for transaction costs, factor orthogonalization, and separate tests in bull and bear markets.

Keywords: margin trading and short selling, asset pricing, mispricing, investor sentiment, principal component analysis

***Supported by** National Natural Science Foundation of China (72573052); Humanities and Social Sciences Foundation of the Ministry of Education of China (24YJC790159); Major Program of National Social Science Foundation of China (25&ZD133); Hunan Provincial Natural Science Foundation of China (2025JJ50425)

*Corresponding author

Email address: hanruixu@hnu.edu.cn (Hanrui Xu)

1. Introduction

Are margin trading and short selling merely the outcomes of market trading activities, or do they contain valuable asset pricing information that has not yet been fully incorporated into stock prices? Margin trading allows investors to amplify their buying demand through leverage, whereas short selling enables investors with negative information or pessimistic beliefs to express their views through short positions (Chang et al., 2014; Li et al., 2015). Therefore, the relative intensity of margin buying and short selling activities may reflect investors' expectations regarding future stock price movements. In particular, short-selling activities and the constraints imposed on short selling have been shown to be closely related to differences of opinion among investors, informed trading based on negative information, and stock mispricing (Boehme et al., 2006; Li et al., 2014).

If these trading activities primarily reflect investors' informational advantages, the information embedded in such transactions should subsequently be incorporated into stock prices. Existing studies generally find that higher short-selling demand or more active short-selling activities predict lower future returns, suggesting that short sellers may possess superior ability to identify negative information (Asquith et al., 2005; Chang et al., 2014). In contrast, the return predictability of margin trading depends critically on the underlying trading motives. When margin purchases reflect valuable positive information, they may contain positive return-predictive signals. However, when margin trading is driven primarily by excessive optimism and speculative demand, stocks with stronger margin-buying activity may become overvalued and subsequently experience lower returns. Prior research shows that elevated investor sentiment and short-sale constraints allow optimistic investors' valuations to dominate market prices, leading to stock overvaluation and lower subsequent returns (Baker and Wurgler, 2006; Gu et al., 2011; Stambaugh et al., 2012). Therefore, identifying whether informed trading or sentiment-driven speculative demand dominates margin and short-selling activities in the Chinese stock market constitutes the central question of this study.

Since its pilot launch in 2010, China's margin trading and securities lending system has undergone continuous expansion in eligible stocks and gradually evolved into a regular trading mechanism (Chang et al., 2014). The

system allows investors to establish leveraged long positions through margin purchases and to express pessimistic views through short selling. From the perspective of market information efficiency, the short-selling mechanism facilitates the incorporation of negative private information into stock prices (Diamond and Verrecchia, 1987; Chang et al., 2014).

Margin trading and securities lending data are typically disclosed at relatively high frequencies and may capture changes in investors' trading demands more promptly than low-frequency accounting information. Compared with outcome-based measures such as stock prices and trading volume, margin and short-selling activities possess clearer directional economic interpretations (Boehmer and Wu, 2013; Chang et al., 2014). However, due to limited availability of lendable securities, relatively high short-selling costs, and trading rules constraints, short-selling activity in the Chinese market remains substantially smaller than margin trading activity, which limits the ability of short selling to fully convey pessimistic information. Therefore, the coexistence of margin trading and short selling, together with their pronounced asymmetry in scale and institutional constraints, provides a unique setting for examining whether margin-related trading information predicts future stock returns and identifying the underlying economic mechanisms.

Existing studies on margin trading and short selling have primarily examined how the introduction of these mechanisms affects stock price efficiency, return distributions, and stock price volatility (Xu and Chen, 2012; Xiao and Kong, 2014). The short-sale constraint theory suggests that when investors hold heterogeneous beliefs and short selling is restricted, pessimistic investors are unable to fully incorporate negative information into prices through trading, allowing optimistic investors to dominate price formation and leading to stock overvaluation (Miller, 1977; Chen et al., 2002). Consistent with this view, prior research shows that short selling facilitates the incorporation of negative information into stock prices and improves market information efficiency (Boehmer and Wu, 2013; Chang et al., 2014).

However, unlike the information discovery role of short selling, margin trading may amplify optimistic demand through leverage and exhibit stronger procyclical characteristics. Given the substantial dominance of margin trading over short selling in the Chinese market, margin purchases and subsequent deleveraging pressure during market downturns may further increase the risk of stock price crashes (Chu and Fang, 2016). Therefore, the effects of margin trading and short selling on stock pricing are unlikely to be symmetric; instead, the two mechanisms may affect stock prices through distinct informational

and behavioral channels.

However, the existing literature has largely treated margin trading and short selling as an institutional arrangement or a quasi-natural experiment, focusing primarily on changes in market outcomes before and after stocks become eligible for margin trading and short selling. Less attention has been paid to whether trading information contained in margin financing and short-selling activities themselves can generate stable cross-sectional return predictability. Although some studies examine the relation between margin balances, short-selling balances, or short-selling activities and future stock returns, they typically rely on individual measures.

While single indicators provide intuitive interpretations, they may fail to capture the multidimensional nature of margin trading and short-selling activities. The absolute magnitude of outstanding balances can be mechanically affected by firm size, while balance-based measures mainly reflect accumulated historical trading positions and may not fully capture marginal changes in trading intensity and investor sentiment. Therefore, it remains unclear whether multidimensional information from margin financing and short selling can be effectively integrated into a composite factor with clear economic interpretation, and whether such a factor contains incremental pricing information beyond traditional firm characteristics.

Our study is also closely related to the literature on investor sentiment and mispricing. Behavioral finance theories suggest that, in the presence of irrational investors and limited arbitrage, investor sentiment can drive stock prices away from their fundamental values, while arbitrage constraints prevent rational investors from eliminating such mispricing in a timely manner, thereby generating predictable differences in future returns (De Long et al., 1990; Barberis et al., 1998). Baker and Wurgler (2006) further show that investor sentiment has stronger effects on stocks that are more difficult to value and subject to greater arbitrage constraints. Stambaugh et al. (2012) find that various asset pricing anomalies become more pronounced during high-sentiment periods, with abnormal returns primarily arising from subsequent corrections of overvalued stocks that are driven by excessive optimism.

Evidence from the Chinese stock market is consistent with this behavioral explanation. Given the high participation of individual investors and relatively limited arbitrage mechanisms in the A-share market, investor sentiment is significantly negatively associated with future short-term stock returns (He et al., 2021). In addition, the lottery preference literature suggests that investors tend to overvalue stocks with lottery-like characteristics and are

willing to pay premiums for opportunities with extremely positive payoffs, causing these stocks to become overvalued and subsequently generate lower future returns (Kumar, 2009; Bali et al., 2011). Therefore, an important question examined in this study is whether the return predictability embedded in margin trading and short-selling information reflects compensation for systematic risks, or instead captures sentiment-driven speculative demand, mispricing, and subsequent price corrections associated with margin trading activities.

To address these questions, we use margin trading and securities lending data for Chinese A-share listed firms from 2010 to 2025 and construct multidimensional measures from both the margin-financing and short-selling sides. On the margin-financing side, we consider changes in margin balances, changes in the ratio of margin balances to tradable market capitalization, and a margin sentiment momentum measure. Corresponding measures are constructed for the short-selling side. These indicators capture different dimensions of margin trading and short-selling activities, including marginal changes in trading scale, trading intensity relative to firm size, and trading enthusiasm relative to historical levels.

To account for the potential correlation among indicators within the same trading direction, we perform cross-sectional principal component analysis (PCA) separately for the margin-financing and short-selling measures in each period. We extract the first principal component from each side and combine them after aligning their economic directions to construct a net margin–short composite factor, denoted as **net_PCA**. A higher value of this factor indicates stronger margin-financing activity relative to short-selling activity, reflecting more optimistic aggregate investor expectations.

In our empirical analysis, we first employ portfolio sorting tests to examine the return predictability of the composite factor at daily, weekly, and monthly frequencies, constructing portfolios using both equal weighting and value weighting schemes. We then evaluate whether the returns of the long–short portfolios can be explained by conventional risk exposures using the CAPM, the Fama–French three-factor model, and the five-factor model. In addition, we examine the persistence and incremental predictive power of the factor through holding-period return tests and Fama–MacBeth cross-sectional regressions.

To investigate the underlying economic mechanisms, we further conduct double-sorting analyses based on the composite factor and two behavioral measures: an investor sentiment indicator constructed from online stock forum

comments and the MAX lottery preference measure. Finally, we perform a series of robustness checks from the perspectives of transaction costs, factor orthogonalization, and market conditions by distinguishing between bull and bear markets.

Our first main finding is that the margin–short composite factor significantly and negatively predicts future stock returns. Sorting stocks into ten portfolios based on factor values, we find that portfolios with lower factor values consistently outperform those with higher factor values in subsequent periods. In the daily equal-weighted portfolio results, the Low–High spread portfolio earns an average daily return of 0.114% with a Newey–West adjusted t-statistic of 10.493, corresponding to an annualized return of 28.801%. The results remain qualitatively similar under weekly frequencies and value-weighted portfolios, suggesting that the return premium is not solely driven by small-cap stocks. However, the predictive power becomes substantially weaker at the monthly frequency, indicating that the pricing information embedded in margin trading and short-selling activities is primarily reflected through short-term price adjustments.

The above findings are difficult to reconcile with a pure information advantage explanation. According to the informed trading literature, investors’ trading activities may contain private information that has not yet been fully incorporated into prices, and such information is gradually reflected in asset prices through the trading process (Kyle, 1985; Glosten and Milgrom, 1985). Therefore, if margin trading and short selling primarily capture investors’ informational advantages, stocks with stronger margin-financing activity relative to short selling should be expected to earn higher subsequent returns. However, our empirical results show that stocks with higher factor values instead generate lower future returns.

This evidence suggests that when margin-financing activity dominates short-selling activity, optimistic expectations may have already been incorporated into stock prices through excessive trading demand, causing prices to deviate from fundamentals and subsequently reverse. This interpretation is consistent with theories of investor sentiment and mispricing (De Long et al., 1990; Baker and Wurgler, 2006). Moreover, the predictive relationship is primarily observed at daily and weekly horizons and weakens as the holding period extends, which is more consistent with short-term mispricing correction.

Our second main finding is that the return predictability of the margin–short composite factor cannot be fully explained by conventional risk factors

or commonly used firm characteristics. The Low–High portfolio generates significantly positive alphas under the CAPM, the Fama–French three-factor model, and the five-factor model. For daily portfolios, the estimated alpha is approximately 0.119% per day across all three models, with t-statistics exceeding 10, and the weekly portfolios also deliver significantly positive risk-adjusted returns. These results suggest that conventional risk factors have limited explanatory power for the returns of the long–short portfolio, indicating that the margin–short composite factor does not simply capture exposures to market, size, value, profitability, or investment factors, but instead contains incremental return-predictive information (Fama and French, 1993, 2015).

The Fama–MacBeth cross-sectional regressions further support this conclusion. After controlling for common stock characteristics, including firm size, short-term reversal, turnover, and Amihud illiquidity, the coefficient on the margin–short composite factor remains significantly negative. These results indicate that the return-predictive information contained in the composite factor is not subsumed by known firm characteristics, reversal effects, or liquidity differences, but instead represents independent incremental cross-sectional return predictability.

Further holding-period analyses reveal the time decay pattern of the return predictability embedded in margin trading and short-selling information. Across holding horizons of 1, 5, 10, 20, and 30 trading days, the Low–High portfolio continues to generate significantly positive cumulative returns. However, both return spreads and risk-adjusted performance measures gradually decline as the holding horizon extends. Consistent with these results, Fama–MacBeth regressions based on future cumulative returns show that the predictive power of the composite factor weakens over longer horizons.

These findings suggest that the margin–short composite factor captures short-term price deviations arising from investor trading behavior and delayed information incorporation, which gradually dissipate as arbitrage forces and information diffusion correct the mispricing. This evidence is consistent with the mispricing and limits-to-arbitrage theories (Shleifer and Vishny, 1997).

Our third main finding is that the return predictability of the margin–short composite factor is consistent with a short-term mispricing mechanism driven by investor sentiment and lottery preference. Using an investor sentiment measure constructed from online stock forum comments, we conduct double-sorting analyses and find that the composite factor retains predictive power across different sentiment conditions. However, the Low–High portfolio

returns increase substantially with sentiment intensity, with the long–short portfolio earning an average daily return of 0.079% (t -statistic = 7.80) in the high-sentiment group. Further sorting by sentiment direction shows that the return spread also increases as investor sentiment shifts from pessimistic to optimistic.

These findings suggest that when market sentiment is stronger or investors hold more optimistic expectations, sentiment-driven demand and speculative trading associated with margin financing are more likely to push stock prices away from fundamentals, leading to subsequent return reversals during price corrections (De Long et al., 1990).

Furthermore, we construct the MAX measure based on the maximum daily return over the previous 20 trading days to capture lottery-like characteristics of stocks and investors’ preference for extreme positive return opportunities (Kumar, 2009; Bali et al., 2011). The double-sorting results show that the long–short returns of the composite factor are significantly positive across low-, medium-, and high-MAX portfolios, with the strongest performance observed among high-MAX stocks.

These findings suggest that the changes in investor expectations embedded in margin trading and short-selling information are more likely to translate into price deviations and subsequent return adjustments for stocks with stronger lottery-like characteristics. Because high-MAX stocks tend to attract investors seeking extreme payoff opportunities, the coexistence of optimistic expectations reflected in margin trading and lottery-driven speculative demand may further increase the likelihood of stock overvaluation, thereby enhancing the return predictability of the composite factor.

Our main findings remain robust across a series of additional tests. First, after incorporating transaction costs of 1, 2, and 5 basis points, the majority of daily and weekly long–short portfolio returns remain positive, with the weekly portfolios exhibiting particularly strong robustness across different cost assumptions. This evidence suggests that the return premium is not solely driven by an unrealistic zero-transaction-cost assumption.

Second, after orthogonalizing the composite factor with respect to firm size and short-term reversal, as well as turnover and illiquidity, the Fama–MacBeth coefficients of the orthogonalized factor remain significantly negative. Finally, the Low–High portfolio continues to generate positive returns in both bull and bear markets, indicating that the return predictability of the factor is not dependent on a particular market regime.

This paper makes contributions to the literature in three ways. First,

it extends the literature on margin trading, short selling, and asset pricing. Existing studies primarily treat margin trading and short selling as institutional arrangements and examine their effects on market liquidity, volatility, information efficiency, and corporate decisions. In contrast, we directly investigate the asset pricing implications of margin trading and short-selling activities themselves. We show that margin–short trading information not only reflects changes in market conditions but also contains persistent cross-sectional return-predictive signals, thereby extending the study of margin trading beyond institutional effects to the asset pricing domain.

Second, this paper provides a new empirical design for constructing asset pricing factors from multidimensional trading behavior information. Unlike existing studies that rely on individual measures such as margin balances, short-selling balances, or trading volume, we construct a set of indicators from both the margin-financing and short-selling sides, including changes in trading activity, relative trading intensity, and sentiment momentum measures. We then apply principal component analysis to extract the common information within each trading direction and construct a net composite factor that captures the relative strength between long and short trading forces.

This approach preserves the distinct economic meanings of margin financing and short selling while reducing information redundancy across multiple indicators, providing a unified framework for integrating high-frequency trading variables that exhibit opposite economic directions and substantial correlations.

Third, this paper contributes to the literature on investor sentiment, lottery preference, and stock mispricing. We show that the return predictability of the margin–short composite factor exhibits a strong short-term pattern and becomes more pronounced among stocks with higher sentiment, more optimistic investor expectations, and stronger lottery-like characteristics. These findings suggest that margin trading and short-selling information not only reflects investors’ rational expectations about future returns, but also captures price deviations arising from leverage-driven demand, sentiment-based trading, and speculative preferences. Therefore, this paper provides new empirical evidence on how investor irrationality enters stock prices through margin trading and short selling and generates subsequent return reversals.

The remainder of the paper is organized as follows. Section 2 describes the sample and data sources and presents the construction of the variables and the margin–short composite factor. Section 3 reports the results of the portfolio sorting tests, risk-adjusted return analyses, holding-period tests,

and Fama–MacBeth cross-sectional regressions. Section 4 examines the economic mechanisms related to investor sentiment and lottery preference. Section 5 presents robustness tests accounting for transaction costs, factor orthogonalization, and different market conditions. Section 6 concludes.

2. Data and Variable Construction

2.1. Sample Selection, Data Sources, and Institutional Background

2.1.1. Sample Selection and Data Sources

We study A-share firms listed on the Shanghai Stock Exchange and the Shenzhen Stock Exchange. Given that China’s margin trading and securities lending program was launched on a pilot basis in 2010 and subsequently expanded and refined over time, our sample covers daily, weekly, and monthly observations from April 19, 2010 to December 31, 2025. This period spans several important episodes in the Chinese stock market, including the severe market turmoil in 2015, and thus provides broad coverage of market dynamics over the evolution of the margin trading and securities lending system. The sample is therefore well suited to capturing the characteristics of China’s equity market under this evolving institutional framework.

Data on margin trading and short selling, including margin balances, short-selling balances, and other related trading measures, are primarily obtained from the Wind database. Stock return and firm-level characteristic data, including stock returns, daily closing prices, market capitalization, and indicators identifying Special Treatment (ST) and Particular Transfer (PT) stocks, are collected from the Wind and CSMAR databases. Asset pricing factor data used in the CAPM and the Fama–French three- and five-factor models (FF3 and FF5), together with the risk-free rate, are obtained from the CSMAR database.

To ensure data quality, we apply the following screening procedures to the raw sample. First, we exclude stocks designated as ST or *ST, as well as stocks in the delisting consolidation period, to avoid potential contamination from special trading conditions. Second, missing values in margin trading and short-selling variables are treated according to their economic interpretation. Observations prior to a stock becoming eligible for margin trading and short selling are removed. After a stock becomes eligible, observations for which both the margin balance and short-selling balance are missing are treated as having no valid margin or short-selling trading record on that trading day and are therefore excluded. If only one of the two balances is missing,

the missing value is set to zero. Third, to mitigate the influence of extreme observations, all continuous variables are winsorized at the 1st and 99th percentiles. After these screening and data-cleaning procedures, the final sample consists of 5,298,413 daily observations, 1,092,880 weekly observations, and 256,599 monthly observations over the period from April 19, 2010 to December 31, 2025. These observations form the basis for the subsequent construction of margin trading and short-selling measures, portfolio sorting analyses, and cross-sectional regressions.

2.1.2. Evolution of the Eligible Stock Universe

To document the implementation of China’s margin trading and securities lending system and the evolution of its market coverage, we track changes in the number of stocks eligible for margin trading and short selling over the sample period.



Figure 1: Evolution of the Margin Trading and Short-Selling Eligible Stock Universe, 2010–2025

Figure 1 plots the annual evolution in the number of stocks eligible for margin trading and short selling in China from 2010 to 2025. The bars report the total number of margin-eligible and short-selling-eligible stocks in each year, while the lines indicate annual additions to the eligible stock universe.

The figure reveals a sustained expansion in the eligible stock universe over the sample period, with margin and short-selling eligibility expanding largely in parallel.

The expansion of the eligible stock universe was not uniform over time but evolved through several distinct stages. From 2010 to 2012, the system remained in its initial pilot phase, with a relatively limited number of eligible stocks and only modest market coverage. After 2013, the program entered a period of rapid expansion, accompanied by a substantial increase in the number of eligible stocks. Since 2023, the pace of new additions has gradually stabilized, indicating that margin trading and short selling have become an increasingly established component of China’s regular market trading mechanism.

A closer examination reveals pronounced increases in newly eligible stocks in 2013, 2019, and 2022, each coinciding with major regulatory expansions of China’s margin trading and securities lending system. In January 2013, the Shanghai Stock Exchange issued an expansion notice, marking the beginning of a more systematic enlargement of the eligible universe. In August 2019, under the guidance of the CSRC, the Shanghai and Shenzhen stock exchanges revised their implementation rules, expanding the eligible stock universe to 1,600 stocks and refining related collateral arrangements. In October 2022, the two exchanges announced another round of expansion, after which the system entered a more stable phase of development.

The gradual expansion of the margin trading and securities lending system provides the empirical basis for our subsequent factor tests and a useful institutional setting for examining the asset pricing implications of margin and short-selling information across different market conditions.

2.2. Construction of Margin and Short-Selling Measures

Margin and short-selling balances provide a broad measure of the overall scale of margin trading and short-selling activities. However, directly using the difference between margin and short-selling balances as the main explanatory variable has several limitations. First, listed firms differ substantially in market capitalization, and the absolute levels of margin and short-selling balances are therefore mechanically related to firm size, which weakens their comparability across firms in the cross section (Wan, 2024). Second, both margin and short-selling balances are stock variables that primarily capture the cumulative outcome of past trading activity and therefore may not adequately reflect marginal changes in investor sentiment, market expectations, or trading

intensity. Accordingly, rather than relying on the raw difference between margin and short-selling balances, we construct separate measures for the two trading directions along three dimensions—balance growth, relative trading intensity, and sentiment momentum—to improve cross-sectional comparability and capture richer time-varying information. These measures form the basis for the subsequent principal component analysis and construction of the composite factor.

Table 1: Construction of Margin Trading and Short-Selling Measures

Measure	Formula	Interpretation	Signal Direction
Change in Margin Balance (<i>fin_bal_chg</i>)	$\frac{FinBal_{i,t} - FinBal_{i,t-1}}{FinBal_{i,t-1}}$	Measures the period-over-period growth in the amount of margin financing and captures whether investors are increasing leveraged long positions over the short run.	Optimistic; a higher value indicates stronger bullish sentiment
Change in Relative Margin Balance (<i>fin_ratio_chg</i>)	$\frac{\frac{FinBal_{i,t}}{MktCap_{i,t}} - \frac{FinBal_{i,t-1}}{MktCap_{i,t-1}}}{\frac{FinBal_{i,t-1}}{MktCap_{i,t-1}}}$	Scales the margin balance by tradable market capitalization to mitigate firm-size effects and captures changes in relative margin-financing intensity.	Optimistic; a higher value indicates stronger bullish sentiment
Margin Sentiment Momentum(<i>fin_sent_mom</i>)	$\frac{FinBal_{i,t}}{\frac{1}{4} \sum_{k=1}^4 FinBal_{i,t-k}}$	Compares the current margin balance with its average over the previous four periods and captures whether margin-financing activity is unusually strong relative to its recent history.	Values above 1 indicate elevated margin-financing sentiment
Change in Short-Selling Balance (<i>short_bal_chg</i>)	$\frac{ShortBal_{i,t} - ShortBal_{i,t-1}}{ShortBal_{i,t-1}}$	Measures the period-over-period growth in short-selling activity and captures whether investors are increasing bearish positions.	Pessimistic; a higher value indicates stronger bearish sentiment
Change in Relative Short-Selling Balance(<i>short_ratio_chg</i>)	$\frac{\frac{ShortBal_{i,t}}{MktCap_{i,t}} - \frac{ShortBal_{i,t-1}}{MktCap_{i,t-1}}}{\frac{ShortBal_{i,t-1}}{MktCap_{i,t-1}}}$	Scales the short-selling balance by tradable market capitalization to mitigate firm-size effects and captures changes in relative short-selling intensity.	Pessimistic; a higher value indicates stronger bearish sentiment
Short-Selling Sentiment Momentum(<i>short_sent_mom</i>)	$\frac{ShortBal_{i,t}}{\frac{1}{4} \sum_{k=1}^4 ShortBal_{i,t-k}}$	Compares the current short-selling balance with its average over the previous four periods and captures whether short-selling activity is unusually strong relative to its recent history.	Values above 1 indicate elevated short-selling intensity

Margin trading typically reflects leveraged buying by investors and therefore contains information about market sentiment and risk-taking preferences (Seguin, 1990). To capture margin-financing activity more comprehensively, we construct three measures that characterize balance growth, relative trading intensity, and sentiment momentum: the change in margin balance, the change in relative margin balance, and margin sentiment momentum.

Consistent with the construction of the margin-financing measures, we construct three corresponding measures on the short-selling side: the change in short-selling balance, the change in relative short-selling balance, and short-selling sentiment momentum. Short selling provides investors with a trading channel through which to express negative expectations and private information, and prior studies generally show that short-selling activity contains information about future returns and contributes to price discovery (Diamond and Verrecchia, 1987; Goyal et al., 2025). The short-selling measures therefore provide a complementary perspective on the information embedded in margin trading and short-selling activities.

In summary, we construct multidimensional measures for both margin-financing and short-selling activities, which form the basis for the subsequent extraction of the composite factor.

2.3. Construction of the Margin–Short Composite Factor

Although the measures described above capture different dimensions of margin trading and short-selling activity, they may contain substantial overlapping information and exhibit high cross-correlations. Directly including them in empirical tests may therefore introduce multicollinearity and obscure their common information content. To extract the common components of the margin-financing and short-selling measures, we employ principal component analysis (PCA) to reduce dimensionality and synthesize the underlying signals (Bai and Ng, 2002; Kelly et al., 2019).

Because margin trading and short selling reflect investors’ optimistic and pessimistic expectations, respectively, the two sides contain distinct information while jointly capturing the overall information embedded in margin trading and short-selling activity (Wan, 2024). Accordingly, we first extract principal components separately from the margin-financing and short-selling sides and then combine them to construct a net factor, thereby improving its economic interpretability. Specifically, the margin-financing block includes the change in margin balance(*fin_bal_chg*), the change in relative margin balance(*fin_bal_chg*), and margin sentiment

momentum(*fin_sent_mom*), whereas the short-selling block includes the change in short-selling balance(*short_bal_chg*), the change in relative short-selling balance(*short_ratio_chg*), and short-selling sentiment momentum(*short_sent_mom*). These variables are used as inputs for the two separate PCA procedures.

First, within the cross section of each trading period, we winsorize each of the six underlying measures at the 1st and 99th percentiles to mitigate the influence of extreme observations. We then standardize each variable cross-sectionally to have a mean of zero and a standard deviation of one. This procedure ensures comparability across measures and prevents the principal component analysis from being dominated by variables with larger numerical scales.

After completing the cross-sectional preprocessing, we conduct principal component analysis separately for the margin-financing and short-selling sides. Specifically, within each cross section, we extract the first principal component from the three standardized margin-financing measures and denote it as fin_{PCA} . We similarly extract the first principal component from the three standardized short-selling measures and denote it as $\text{short}_{\text{PCA}}$. Because the first principal component explains the largest possible share of the common variation among the original variables over all linear combinations, it can be interpreted as a summary measure of trading characteristics on the margin-financing or short-selling side. The composite factor for the margin-financing side can therefore be expressed as:

$$\text{fin_PCA} = \text{PC1}(\text{fin_bal_chg}, \text{fin_ratio_chg}, \text{fin_sent_mom}) \quad (1)$$

The composite factor for the short-selling side can be expressed as:

$$\begin{aligned} \text{short_PCA} = \text{PC1}(\text{short_bal_chg}, \text{short_ratio_chg}, \\ \text{short_sent_mom}) \end{aligned} \quad (2)$$

To assess whether retaining only the first principal component is appropriate, we examine its explained variance on both the margin-financing and short-selling sides. Across the daily, weekly, and monthly samples, the first principal component explains approximately 80% of the variation in the margin-financing measures and more than 84% of the variation in the short-selling measures on average. This high explanatory power indicates

that the first principal component captures most of the common information contained in the underlying measures. We therefore retain it as the summary measure of the core trading characteristics on each side.

Table 2: Variance Explained by the First Principal Component across Frequencies

Frequency	Average Variance Explained by Margin-Side PC1	Average Variance Explained by Short-Selling-Side PC1
Daily	80.2%	86.1%
Weekly	80.8%	84.9%
Monthly	81.0%	84.4%

Because the sign of a principal component is inherently indeterminate, the economic interpretation of high and low principal-component scores may reverse across dates if the direction is not standardized. To ensure consistent economic interpretation over time, we align the signs of the first principal components on both the margin-financing and short-selling sides. For the margin-financing side, we use the change in relative margin balance (*fin_ratio_chg*) as the directional anchor and compute its correlation with the first principal-component score. If the correlation is negative, we multiply the first principal component by -1 , so that a higher margin-side component consistently indicates stronger margin-financing activity. We apply the same procedure to the short-selling side, using the change in relative short-selling balance (*short_ratio_chg*) as the directional anchor, so that a higher short-selling component consistently indicates stronger short-selling activity.

After obtaining the sign-aligned fin_{PCA} and $short_{PCA}$, we further construct the net margin–short composite factor, *net_PCA*, to capture the relative strength of margin-financing and short-selling activities. The net factor is defined as:

$$net_PCA = fin_PCA - short_PCA \quad (3)$$

A higher value of *net_PCA* indicates that margin-financing activity is stronger relative to short-selling activity, reflecting more optimistic aggregate investor expectations. Conversely, a lower value of *net_PCA* indicates relatively stronger short-selling activity and therefore more pessimistic expectations. By construction, *net_PCA* preserves the multidimensional information contained in both margin-financing and short-selling activities while

compressing overlapping information through principal component analysis, thereby providing a parsimonious core explanatory variable for the subsequent portfolio sorting tests and cross-sectional asset pricing analyses.

2.4. Descriptive Statistics

Table 3: Descriptive Statistics for the Daily Sample

Variable	N	Mean	SD	P25	Median	P75
<i>fin_bal_chg</i>	5298411	-0.001	0.996	-0.414	-0.07	0.311
<i>fin_ratio_chg</i>	5298411	-0.001	0.997	-0.46	-0.028	0.422
<i>fin_sent_mom</i>	5298411	0	1	-0.461	-0.105	0.315
<i>short_bal_chg</i>	4382866	0	1	-0.257	-0.112	0.094
<i>short_ratio_chg</i>	4382866	0	1	-0.237	-0.1	0.081
<i>short_sent_mom</i>	4449051	0	1	-0.325	-0.085	0.191
<i>fin</i> _{PCA}	5298411	0	1.542	-0.688	-0.116	0.521
<i>short</i> _{PCA}	4382866	0	1.580	-0.451	-0.169	0.194
<i>net_PCA</i>	5298413	0	1	-0.375	0.005	0.404

Table 3 reports descriptive statistics for the six underlying margin trading and short-selling measures, *fin*_{PCA}, *short*_{PCA}, and the composite factor *net_PCA* in the daily sample. After winsorization and cross-sectional standardization, the six underlying measures have means close to zero and standard deviations close to one, confirming that they are placed on a common scale and are therefore directly comparable in the subsequent principal component analysis.

The percentile statistics indicate meaningful cross-sectional variation in both the margin-financing and short-selling measures, providing sufficient dispersion for extracting common components. The means of *fin*_{PCA} and *short*_{PCA} are both close to zero, with standard deviations of 1.542 and 1.580, respectively, suggesting that the principal components retain substantial cross-sectional information from the underlying measures. The composite factor *net_PCA* has a mean of zero and a standard deviation of one, with the 25th, 50th, and 75th percentiles of -0.375, 0.005, and 0.404, respectively. Overall, *net_PCA* is centered around zero and exhibits sufficient cross-sectional dispersion for subsequent portfolio sorting and asset pricing tests.

Overall, the variables exhibit distributional properties typical of financial market data. The underlying measures capture meaningful cross-sectional heterogeneity in margin trading and short-selling activity, while the composite factor displays sufficient dispersion for the subsequent portfolio sorting and regression analyses.

3. Quantitative Trading Strategies Based on Margin Trading and Short-Selling Information of Listed Firms

3.1. Portfolio Sorting Tests

3.1.1. Portfolio Return Analysis

To evaluate the feasibility of an investment strategy based on the `net_PCA` factor, we employ univariate portfolio sorting. Portfolio sorting is widely used in asset pricing research to examine the relation between stock characteristics and future returns (Fama and French, 1992, 1993). Specifically, at each portfolio formation date, we sort stocks into ten portfolios, labeled D1 through D10, based on the cross-sectional distribution of `net_PCA` from low to high. D1 contains stocks with the lowest factor values, whereas D10 contains those with the highest factor values. We then match the portfolio assignments at the formation date with individual stock returns in the subsequent period, ensuring that the sorting procedure uses only information available at the time of portfolio formation and thereby avoiding look-ahead bias.

Portfolio returns are calculated using both equal weighting and tradable-market-capitalization weighting. Equal-weighted returns capture the average performance of the stocks in each portfolio, whereas value-weighted returns reduce the influence of small-cap stocks and allow us to assess whether the sorting results are primarily driven by smaller firms. We further construct a Low–High long–short portfolio by taking a long position in D1, which contains stocks with the lowest factor values, and a short position in D10, which contains stocks with the highest factor values. The statistical significance of average portfolio returns is evaluated using Newey–West heteroskedasticity- and autocorrelation-consistent standard errors (Newey and West, 1987).

We conduct portfolio sorts at daily, weekly, and monthly frequencies to examine whether the return predictability of `net_PCA` is primarily short-lived or persists over longer horizons.

Turning to the equal-weighted results, the Low–High portfolio earns an average daily return of 0.114%, with a Newey–West t -statistic of 10.493 and

Table 4: Returns of Key Portfolios across Frequencies under Equal Weighting

Frequency	Portfolio	Mean Return (%)	Newey-West t	Annual Return (%)
Daily	<i>D1</i>	0.095	3.283	23.989
	<i>D10</i>	-0.019	-0.65	-4.812
	<i>Low-High</i>	0.114	10.493	28.801
Weekly	<i>D1</i>	0.238	2.080	12.373
	<i>D10</i>	0.008	0.062	0.423
	<i>Low-High</i>	0.230	4.030	11.950
Monthly	<i>D1</i>	0.922	1.827	11.066
	<i>D10</i>	0.641	1.1524	7.688
	<i>Low-High</i>	0.281	1.169	3.377

Table 5: Returns of Key Portfolios across Frequencies under Value Weighting

Frequency	Portfolio	Mean Return (%)	Newey-West t	Annual Return (%)
Daily	<i>D1</i>	0.077	3.162	19.320
	<i>D10</i>	-0.012	-0.463	-2.955
	<i>Low-High</i>	0.088	5.691	22.275
Weekly	<i>D1</i>	0.260	2.575	13.499
	<i>D10</i>	0.010	0.087	0.541
	<i>Low-High</i>	0.249	3.125	12.958
Monthly	<i>D1</i>	0.627	1.506	7.525
	<i>D10</i>	0.518	1.021	6.215
	<i>Low-High</i>	0.109	0.342	1.310

an annualized return of 28.801%. At the weekly frequency, the Low–High portfolio earns an average return of 0.230%, with a t -statistic of 4.030 and an annualized return of 11.950%. At the monthly frequency, however, the average return declines to 0.281%, with a t -statistic of only 1.169, and is no longer statistically significant. Taken together, these results indicate that `net_PCA` has strong short-horizon return predictability, but its predictive power weakens as the investment horizon lengthens.

The value-weighted results are directionally consistent with the equal-weighted results. At the daily and weekly frequencies, the Low–High portfolio earns average returns of 0.088% and 0.249%, with corresponding t -statistics of 5.691 and 3.125, respectively; both are significantly positive. At the monthly frequency, the return is again statistically insignificant. Although the magnitude and statistical significance of the returns are generally weaker under value weighting than under equal weighting, the daily and weekly results remain robust, indicating that the predictive power of `net_PCA` is not entirely driven by small-cap stocks (Stambaugh et al., 2012).

Because `net_PCA` is constructed as $fin_{PCA} - short_{PCA}$, a higher factor value indicates stronger optimistic trading pressure on the margin-financing side relative to the short-selling side. If the market response to this information were persistent, stocks with higher factor values should earn higher subsequent returns. However, our empirical results show that the D1 portfolio consistently outperforms the D10 portfolio, indicating that stronger optimistic trading pressure does not translate into higher future returns but is instead associated with lower subsequent returns. This pattern is consistent with investor overreaction and short-term mispricing: optimistic trading may push current prices above fundamentals, followed by return reversals as prices subsequently correct (De Long et al., 1990; Daniel et al., 1998; Greenwood and Shleifer, 2014). Taken together with the frequency-specific results, the predictive power of `net_PCA` is concentrated primarily at the daily and weekly horizons, which is more consistent with short-term mispricing than with long-run risk compensation.

3.1.2. Cumulative Return Analysis

Following the portfolio sorting results, we further examine whether the strategy generates persistent and stable returns by plotting cumulative return curves for the key portfolios. These curves provide a more intuitive view of how portfolio performance evolves over the full sample period and help distinguish whether the strategy’s returns are concentrated in a few episodic

market phases or instead accumulate persistently over time.

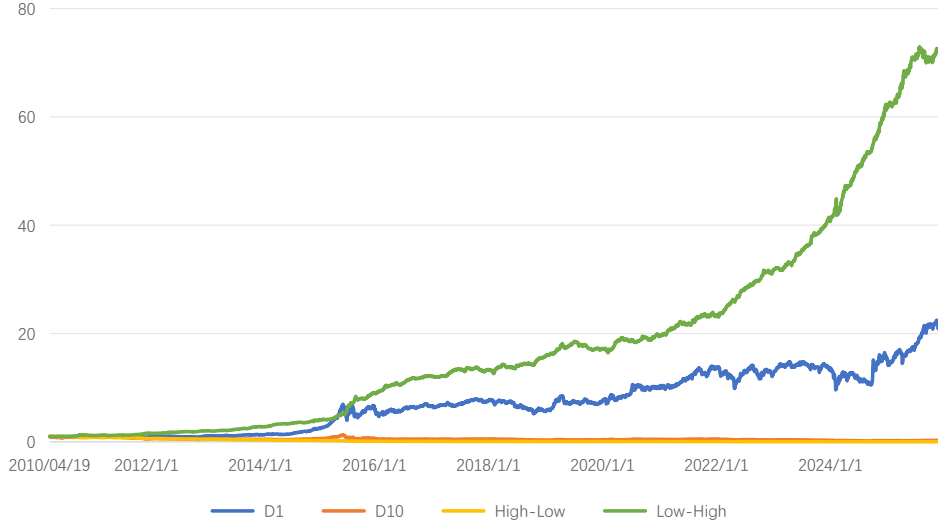


Figure 2: Daily Cumulative Returns under Equal Weighting

Figures 2 through 5 present the cumulative returns of the key portfolios at the daily and weekly frequencies under both equal weighting and value weighting. Overall, the cumulative return curves of the Low–High portfolio exhibit a sustained upward trend and remain substantially above those of the other three portfolios. This pattern indicates that the sorting results documented earlier using average and annualized returns are not confined to isolated periods but display persistent performance over the sample period.

Examining the evolution of cumulative returns over time, all four figures show a pronounced widening of return differentials around 2015. In particular, the cumulative return of the Low–High portfolio rises sharply during this period, while the gap between D1 and D10 also expands substantially. This pattern is consistent with the view that, during periods of heightened market volatility, sentiment signals embedded in margin trading and short-selling activity have stronger implications for subsequent cross-sectional return differences.

Overall, the cumulative return analysis provides further evidence that portfolios constructed on the basis of `net_PCA` exhibit strong profitability. In particular, the Low–High portfolio shows a stable upward cumulative-return

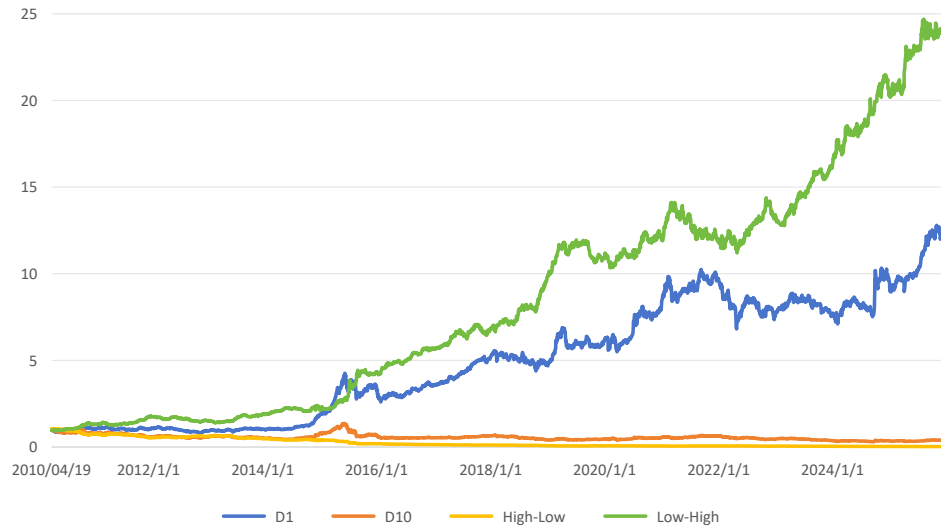


Figure 3: Daily Cumulative Returns under Value Weighting

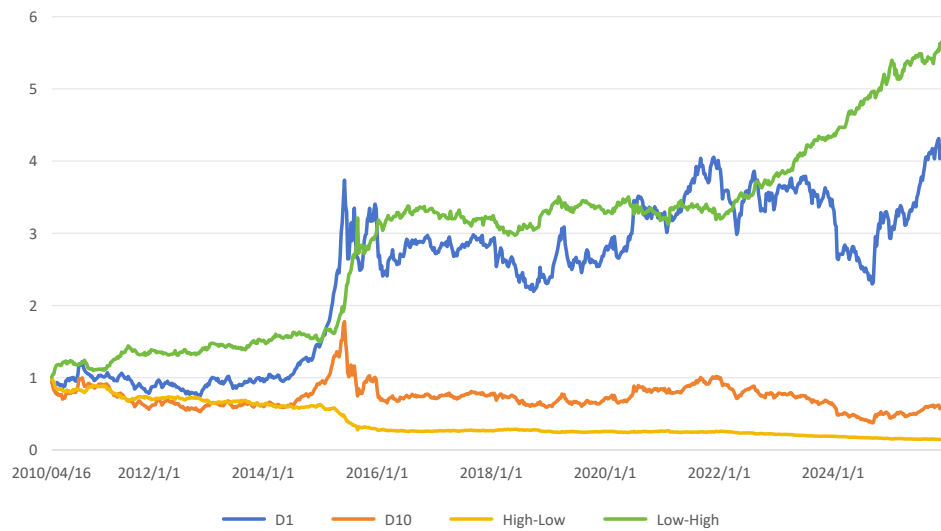


Figure 4: Weekly Cumulative Returns under Equal Weighting

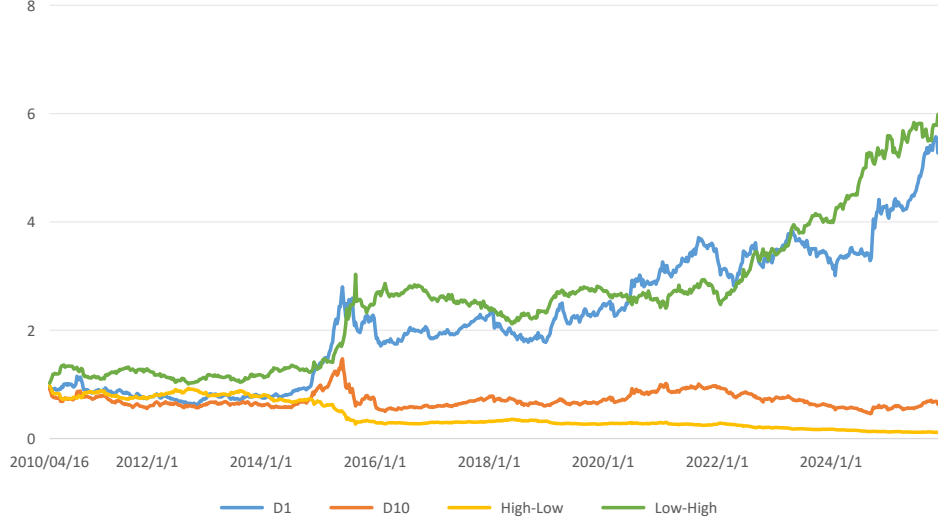


Figure 5: Weekly Cumulative Returns under Value Weighting

pattern at both the daily and weekly frequencies, with returns accumulating more rapidly during periods of pronounced fluctuations in market sentiment. These patterns are consistent with the view that sentiment signals embedded in margin trading and short-selling activity become more strongly associated with subsequent price corrections when market sentiment is more volatile.

3.1.3. Risk–Return Characteristics

After examining the average and cumulative returns of the sorted portfolios, we next evaluate the quantitative trading strategy based on margin trading and short-selling information from a risk–return perspective. We focus on three measures - volatility, the Sharpe ratio, and maximum drawdown - to assess the overall risk–return characteristics of the portfolios across different frequencies.

Table 6 reports the overall risk–return characteristics of the equal-weighted portfolios across different frequencies. Overall, the Low–High portfolio exhibits the most favorable risk–return profile at each frequency, with the lowest annualized volatility, the highest Sharpe ratio, and the smallest maximum drawdown. These results indicate that the long–short strategy based on `net_PCA` not only delivers higher returns but also does so with lower return

Table 6: Risk–Return Measures of Key Portfolios across Frequencies under Equal Weighting

Frequency	Portfolio	Annual Volatility (%)	<i>Sharpe</i> <i>Ratio</i>	Max Drawdown (%)
Daily	<i>D1</i>	26.156	0.917	42.961
	<i>D10</i>	27.761	-0.173	86.756
	<i>High-Low</i>	10.121	-2.846	98.903
	<i>Low-High</i>	10.121	2.846	11.586
Weekly	<i>D1</i>	23.181	0.534	43.316
	<i>D10</i>	26.472	0.016	78.901
	<i>High-Low</i>	11.559	-1.034	85.597
	<i>Low-High</i>	11.559	1.034	15.368
Monthly	<i>D1</i>	23.913	0.463	46.529
	<i>D10</i>	26.338	0.292	55.720
	<i>High-Low</i>	11.406	-0.296	68.366
	<i>Low-High</i>	11.406	0.296	45.312

volatility and smaller extreme losses.

Across frequencies, the risk–return advantage of the Low–High portfolio is most pronounced at the daily frequency. Its annualized volatility is only 10.121%, substantially lower than that of the other portfolios, while its Sharpe ratio reaches 2.846 and its maximum drawdown is only 11.586%, far below that of the one-sided portfolios and the 98.903% maximum drawdown of the High–Low portfolio. These results indicate that, at the daily frequency, the Low–High strategy exhibits favorable risk-control characteristics. The weekly results are broadly consistent with this pattern, suggesting that the strategy remains robust over relatively short holding horizons. At the monthly frequency, however, the Sharpe ratio of the Low–High portfolio declines to 0.296, while its maximum drawdown rises to 45.312%, indicating substantially weaker risk–return performance than at the daily and weekly frequencies. Taken together, these findings further suggest that the effectiveness of the net_PCA-based strategy is concentrated at shorter holding horizons; as the horizon lengthens, both its return advantage and risk-control benefits diminish.

3.2. Risk-Adjusted Return Tests

To determine whether the portfolios generate genuine abnormal returns rather than merely reflecting exposures to conventional risk factors such as market, size, and value, we evaluate portfolio performance using the CAPM, the Fama–French three-factor model, and the Fama–French five-factor model. If the portfolio intercept, α , remains significantly positive after controlling for these risk factors, the strategy’s returns cannot be fully explained by standard asset pricing factors, providing evidence that the quantitative trading strategy based on margin trading and short-selling information delivers significant risk-adjusted excess returns.

Specifically, the CAPM is specified as:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + \varepsilon_t \quad (4)$$

The Fama–French three-factor model further incorporates the size factor, SMB, and the value factor, HML, and is specified as:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + s_i \times \text{SMB}_t + h_i \times \text{HML}_t + \varepsilon_t \quad (5)$$

Building on this specification, the Fama–French five-factor model additionally includes the profitability factor, RMW, and the investment factor, CMA, and is specified as:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + s_i \times \text{SMB}_t + h_i \times \text{HML}_t + \omega_i \times \text{RMW}_t + c_i \times \text{CMA}_t + \varepsilon_t \quad (6)$$

The intercepts estimated from the CAPM, FF3, and FF5 regressions are denoted by α_{CAPM} , α_{FF3} , and α_{FF5} , respectively, and represent the risk-adjusted abnormal returns of each portfolio under the corresponding asset pricing models.

Table 7: Risk-Adjusted Returns of the Low–High Portfolio

Frequency	Model	<i>Alpha</i> (%)	<i>t(Alpha)</i>	<i>Mkt-RF</i>	<i>SMB</i>	<i>HML</i>	<i>RMW</i>	<i>CMA</i>	<i>R2</i>
Daily	CAPM	0.119***	10.805	-0.030					0.003
	FF3	0.119***	10.838	-0.033	0.020	0.011			0.004
	FF5	0.119***	10.872	-0.035	0.014	0.015	-0.025	0.004	0.004
Weekly	CAPM	0.220***	3.862	-0.034					0.002
	FF3	0.215***	3.770	-0.024	-0.062	-0.010			0.005
	FF5	0.220***	3.889	-0.035	-0.084	0.050	-0.040	-0.161	0.012

At the daily frequency, the estimated α_i of the Low–High portfolio is approximately 0.119% under the CAPM, FF3, and FF5 models, with corresponding t -statistics all exceeding 10 and significance at the 1% level. The α_i estimates vary only marginally across the three specifications, indicating that the portfolio returns are not primarily driven by exposures to the market, size, value, profitability, or investment factors.

The weekly results are consistent with those at the daily frequency. The Low–High portfolio generates α_i estimates ranging from 0.215% to 0.220% across the three models, with corresponding t -statistics between 3.770 and 3.889; all of the estimates are significantly positive. Relative to the daily results, however, the statistical significance of the weekly α_i estimates is weaker, consistent with the earlier portfolio-sorting evidence that the return predictability of net_PCA declines as the investment horizon lengthens.

Taken together, the daily and weekly results show that the Low–High portfolio earns significantly positive α_i under the CAPM, FF3, and FF5 models. The persistence of these alphas across alternative factor specifications

indicates that the return spread cannot be fully explained by conventional risk exposures and represents robust risk-adjusted abnormal performance.

3.3. Holding-Period Return Tests

The preceding results primarily capture one-period-ahead return predictability. To examine whether the strategy remains effective over longer holding horizons, we conduct holding-period tests using daily data and evaluate the performance of the Low–High portfolio across different holding periods, with particular attention to how its returns evolve as the holding horizon lengthens.

Specifically, using `net_PCA` as the sorting variable, we sort the sample stocks into ten portfolios from the lowest to the highest factor values within the cross section of each trading day. We then examine the future cumulative returns of these portfolios over holding horizons of 1, 5, 10, 20, and 30 trading days. For any stock i at time t , the holding-period return over the subsequent h trading days is calculated as the cumulative return over that horizon, that is:

$$R_{i,t+1:t+h} = \prod_{k=1}^h (1 + r_{i,t+k}) - 1 \quad (7)$$

where $r_{i,t+k}$ denotes the daily return of stock i on trading day $t + k$, and h denotes the holding horizon and takes values of 1, 5, 10, 20, and 30 trading days.. For each trading day, we then compute the equal-weighted average of the future h -day cumulative returns of stocks within each portfolio to obtain the portfolio return series for the corresponding holding horizon. Based on these portfolio return series, we construct the Low–High long–short portfolio.

Table 8: Holding-Period Test Results

Horizon	Portfolio	Mean Return (%)	NW t-stat	NW p-value	Annual- ized Return (%)	Sharpe	NW Lag
1d	<i>Low-High</i>	0.114	11.107	1.16E-28	33.365	2.846	1
5d	<i>Low-High</i>	0.208	7.301	2.86E-13	11.014	1.028	4
10d	<i>Low-High</i>	0.310	7.143	9.13E-13	8.115	0.817	9
20d	<i>Low-High</i>	0.438	6.603	4.04E-11	5.666	0.607	19
30d	<i>Low-High</i>	0.431	5.092	3.54E-07	3.678	0.404	29

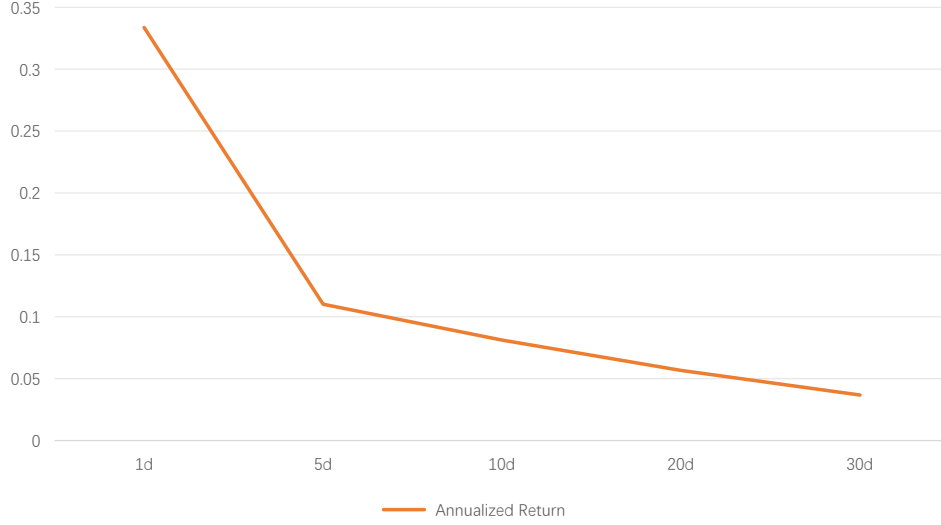


Figure 6: Annualized Returns across Different Holding Periods

Source: Authors' calculations based on Wind, CSMAR, and empirical results.

Overall, the Low–High portfolio earns positive average returns across all holding horizons, with statistically significant Newey–West t -statistics in each case. These results indicate that the portfolio's profitability is not confined to a single holding period but persists, to some extent, over longer holding horizons.

However, from a risk–return perspective, the strategy exhibits a clear deterioration in performance as the holding horizon lengthens. The annualized return declines from 33.365% for the 1-day holding period to 3.678% for the 30-day holding period, while the annualized Sharpe ratio falls steadily from 2.846 to 0.404. These results indicate that although the strategy continues to generate positive returns over longer holding horizons, both the return per unit of time and the risk-adjusted performance decline substantially, with the strongest performance concentrated at shorter holding horizons.

Taken together, the holding-period tests provide further evidence of the persistence of the Low–High strategy's returns, while also revealing a clear pattern of stronger performance at short horizons and weaker performance at longer horizons. This finding further suggests that the factor is unlikely to capture a persistent long-run risk premium. Instead, the weakening of returns

over longer horizons is consistent with gradual correction of sentiment-driven mispricing, as earlier overpricing is progressively reversed and the associated return spread dissipates.

3.4. Fama–MacBeth Cross-Sectional Regressions

3.4.1. Fama–MacBeth Regressions of Future Returns

Portfolio sorting provides an intuitive way to compare return differences across groups formed on the basis of the factor, but it does not characterize the continuous cross-sectional relation between factor values and future returns. We therefore further employ the two-step Fama–MacBeth cross-sectional regression framework to examine the return predictability of `net_PCA` (Fama and MacBeth, 1973). In the first step, at each time t , we run a cross-sectional regression of individual stocks’ next-period returns on contemporaneous `net_PCA` and a set of control variables. The first-stage cross-sectional regression is specified as follows:

$$R_{i,t+1} = \alpha_t + \beta_t \text{net_PCA}_{i,t} + \gamma_t \text{Controls}_{i,t} + \varepsilon_{i,t} \quad (8)$$

where $R_{i,t+1}$ denotes the next-period return of stock i following time t , and $\text{net_PCA}_{i,t}$ denotes the value of the factor for stock i at time t . $\text{Controls}_{i,t}$ denotes the vector of control variables, including firm size (`ln_size`), short-term reversal (`REV`), turnover (`Turnover`), and illiquidity (`ln_Amihud`) (Fama and French, 1992; Datar et al., 1998; Amihud, 2002). The coefficient β_t is the cross-sectional return coefficient on `net_PCA` in period t and captures the relation between the factor and future returns.

In the second step, we average the coefficients estimated from the period-by-period cross-sectional regressions over time and test whether the time-series average coefficient is statistically different from zero. That is:

$$\bar{\beta} = \frac{1}{T} \sum_{t=1}^T \beta_t \quad (9)$$

T denotes the number of periods included in the estimation. If the time-series average coefficient, $\bar{\beta}$, is significantly different from zero, `net_PCA` exhibits a stable cross-sectional relation with next-period returns. In particular, a significantly negative average coefficient indicates that stocks with higher `net_PCA` values subsequently earn lower returns. The regression results are reported below.

Table 9: Fama–MacBeth Regression Results for Future Returns

Variable	Model(1)	Model(2)	Model(3)	Model(4)	Model(5)
<i>Constant</i>	0.04 (1.425)	0.132 (1.112)	0.083 (0.818)	0.458*** (4.463)	0.273*** (2.678)
<i>net_PCA</i>	-0.025*** (-9.430)	-0.025*** (-10.232)	-0.015*** (-6.593)	-0.005** (-2.433)	-0.004** (-2.188)
<i>ln_size</i>		-0.007 (-0.949)	-0.005 (-0.822)	-0.027*** (-4.347)	-0.016** (-2.542)
<i>REV</i>			3.081*** (10.449)	3.832*** (13.183)	3.164*** (10.889)
<i>Turnover</i>				-2.991*** (-9.126)	-2.719*** (-8.245)
<i>ln_Amihud</i>					176.229*** (4.222)

As shown in the table, the coefficient on the core variable *net_PCA* is significantly negative across all model specifications. In Model (1), which includes no control variables, the coefficient on *net_PCA* is -0.025% and is significant at the 1% level. After controlling for firm size (*ln_size*) and short-term reversal (*REV*), the coefficient remains -0.015% and continues to be significant at the 1% level. Given that margin trading and short-selling activity may also be related to trading intensity and liquidity conditions, Models (4) and (5) further include turnover (*Turnover*) and illiquidity (*ln_Amihud*). The results show that, after fully controlling for firm size, short-term reversal, turnover, and illiquidity, the coefficient on *net_PCA* remains negative at -0.004% and is significant at the 5% level. Overall, these results indicate that the margin–short composite factor has robust negative predictive power for future stock returns, and that its predictive content is not fully subsumed by firm size, short-term reversal, or liquidity-related characteristics.

3.4.2. Fama–MacBeth Regressions of Holding-Period Returns

The preceding holding-period tests show that the returns of the Low–High portfolio continue to accumulate over several trading days. In this section, we extend the analysis by conducting Fama–MacBeth regressions using future cumulative holding-period returns as the dependent variable. This approach

allows us to examine, from a cross-sectional perspective, the predictive power of `net_PCA` for future cumulative returns and to assess whether its predictive relation weakens as the holding horizon lengthens.

Table 10: Fama–MacBeth Regression Results for Holding-Period Returns

Horizon	Variable	Mean Coef(%)	t- stat(NW)	Avg N	Avg R2
1d	const	0.039	1.462	1387.543	0.005
1d	<i>net_PCA</i>	-0.025	-9.07	1387.543	0.005
5d	const	0.208	1.848	1385.280	0.005
5d	<i>net_PCA</i>	-0.044	-5.751	1385.280	0.005
10d	const	0.411	1.840	1382.450	0.005
10d	<i>net_PCA</i>	-0.073	-6.157	1382.450	0.005
20d	const	0.829	1.853	1376.774	0.004
20d	<i>net_PCA</i>	-0.113	-5.824	1376.774	0.004
30d	const	1.231	1.859	1371.076	0.004
30d	<i>net_PCA</i>	-0.120	-4.587	1371.076	0.004

Table 10 reports the Fama–MacBeth cross-sectional regression results based on cumulative holding-period returns. Overall, the average coefficient on `net_PCA` is significantly negative across all five holding horizons of 1, 5, 10, 20, and 30 trading days. This evidence indicates that `net_PCA` not only predicts next-period returns but also exhibits a stable cross-sectional relation with future cumulative returns over longer holding horizons.

As the holding horizon increases from 1 to 30 trading days, the corresponding t -statistics decline gradually but remain statistically significant. This pattern suggests that the cross-sectional predictive relation between `net_PCA` and cumulative future returns weakens over longer horizons, while remaining statistically detectable throughout the 30-day window.

Taken together, the portfolio sorting tests, risk-adjusted return analyses, and Fama–MacBeth regressions consistently show that `net_PCA` is significantly negatively associated with future stock returns. This predictive relation remains after controlling for firm size, short-term reversal, turnover, and liquidity characteristics, and is concentrated primarily at shorter forecasting horizons. These findings indicate that `net_PCA` contains incremental cross-sectional return information beyond conventional risk factors and commonly

used stock characteristics. The economic sources of this predictive content, however, require further investigation through the mechanism analyses that follow.

4. Economic Mechanism Analysis

4.1. Framework for Mechanism Analysis

The preceding empirical results show that `net_PCA` has significant predictive power for future stock returns. However, the existing evidence does not fully explain why the factor predicts future returns or why its return predictability becomes markedly stronger during periods of severe market turbulence, such as in 2015. Because margin financing and short selling reflect investors' differing expectations about future market movements, these trading activities may also embed behavioral information related to investor sentiment, risk preferences, and speculative trading. It is therefore necessary to further investigate the economic mechanisms underlying the return predictability of `net_PCA`.

Drawing on behavioral finance theory, we examine two potential mechanisms. First, the investor sentiment and mispricing channel. Stronger investor sentiment may induce overconfidence, herding, and return-chasing behavior, causing stock prices to deviate from fundamental values and generating predictable return differentials as prices subsequently correct. Second, the lottery preference and speculative trading channel. Investors with a stronger preference for extreme positive payoffs are more likely to chase high-MAX stocks, while margin leverage may further amplify such speculative demand, pushing prices above fundamental values in the short run and leading to subsequent return reversals (De Long et al., 1990; Baker and Wurgler, 2006; Bali et al., 2011).

To test these mechanisms, we introduce investor sentiment and MAX measures and conduct double-sorting analyses jointly with `net_PCA`. We then compare the performance of the `net_PCA`-sorted portfolios across different sentiment states and levels of lottery preference. If the Low–High portfolio generates stronger and more statistically significant returns in the high-sentiment or high-MAX groups, this would be consistent with the view that the return predictability of `net_PCA` is related to mispricing and subsequent price correction arising from investor sentiment, lottery preference, and speculative trading.

4.2. Mechanism Analysis Based on Investor Sentiment

4.2.1. Construction of the Investor Sentiment Measure

The investor sentiment data used in this study are primarily obtained from the stock forum comment database compiled by the CNRDS database. Using data from January 1, 2010 to December 31, 2025, we collect the numbers of positive posts, negative posts, and total posts for all listed firms and construct two sentiment measures: sentiment direction and sentiment intensity. The number of positive posts is denoted by Pospostnum , and the number of negative posts is denoted by Negpostnum .

First, we construct a sentiment intensity measure to capture the extent to which investor sentiment deviates from a neutral state. Because both strongly optimistic and strongly pessimistic sentiment indicate a more concentrated market view and increasingly one-sided investor beliefs, we first construct a stock-level sentiment measure based on the ratio of positive to negative posts:

$$\text{Sentiment}_{i,t} = \ln \left(\frac{\text{Pospostnum}_{i,t} + 1}{\text{Negpostnum}_{i,t} + 1} \right) \quad (10)$$

On this basis, we take the absolute value of the sentiment measure to obtain the sentiment intensity measure:

$$\text{Strength}_{i,t} = |\text{Sentiment}_{i,t}| \quad (11)$$

A higher value of the measure indicates a greater deviation of investor sentiment from the neutral state. Whether this reflects stronger optimism or deeper pessimism, a larger value corresponds to more intense market sentiment. Conversely, a lower value indicates relatively muted sentiment conditions.

Second, we retain the sign of the sentiment measure to capture the direction of investor sentiment, that is, whether overall market sentiment is more optimistic or more pessimistic:

$$\text{Sentiment}_{i,t} = \ln \left(\frac{\text{Pospostnum}_{i,t} + 1}{\text{Negpostnum}_{i,t} + 1} \right) \quad (12)$$

When the measure is greater than zero, the number of positive posts exceeds the number of negative posts, indicating that overall investor sentiment is relatively optimistic. Conversely, when the measure is below zero, overall investor sentiment is relatively pessimistic.

In addition, to maintain consistency with the construction of the `net_PCA` factor, we standardize both the sentiment intensity and sentiment direction measures cross-sectionally by trading day. The resulting standardized measures are then used in the subsequent double-sorting tests.

4.2.2. Double-Sorting Tests by Sentiment Intensity

To examine whether the return predictability of the margin-short composite factor `net_PCA` varies with the strength of investor sentiment, we conduct a double-sorting mechanism test. Within the cross section of each trading day, stocks are first sorted into low-, medium-, and high-sentiment-intensity groups according to the sentiment intensity measure $Strength_{i,t}$. Within each sentiment-intensity group, stocks are then further sorted into three portfolios based on `net_PCA`, denoted by PCA1, PCA2, and PCA3. This procedure yields a 3×3 set of double-sorted portfolios.

After forming the portfolios, we continue to use the t -period sorting and $t+1$ -period return-matching procedure to compute the equal-weighted average return of each portfolio in the subsequent period. We then construct, within each sentiment-intensity group, a `net_PCA`-based long-short portfolio. The empirical results are reported below.

Table 11: Double-Sorting Results for `net_PCA` and Sentiment Intensity

	PCA1	PCA2	PCA3	L-H
<i>Strength_L</i>	0.027 (0.95)	0.022 (0.78)	-0.014 (-0.50)	0.041*** (4.75)
<i>Strength_M</i>	0.054* (1.90)	0.026 (0.95)	0.003 (0.11)	0.050*** (5.51)
<i>Strength_H</i>	0.128*** (4.71)	0.068*** (2.63)	0.049* (1.71)	0.079*** (7.80)

As shown in Table 11, `net_PCA` exhibits return predictability across different sentiment-intensity states, although the magnitude of this predictability varies substantially. In the low-sentiment-intensity group, the average returns of the PCA1, PCA2, and PCA3 portfolios are 0.027%, 0.022%, and -0.014%, respectively. The corresponding Low-High portfolio earns an average return of 0.041%, equivalent to an annualized return of 10.83%, indicating that `net_PCA` can still differentiate stocks with different future return outcomes even when market sentiment is relatively muted. As sentiment

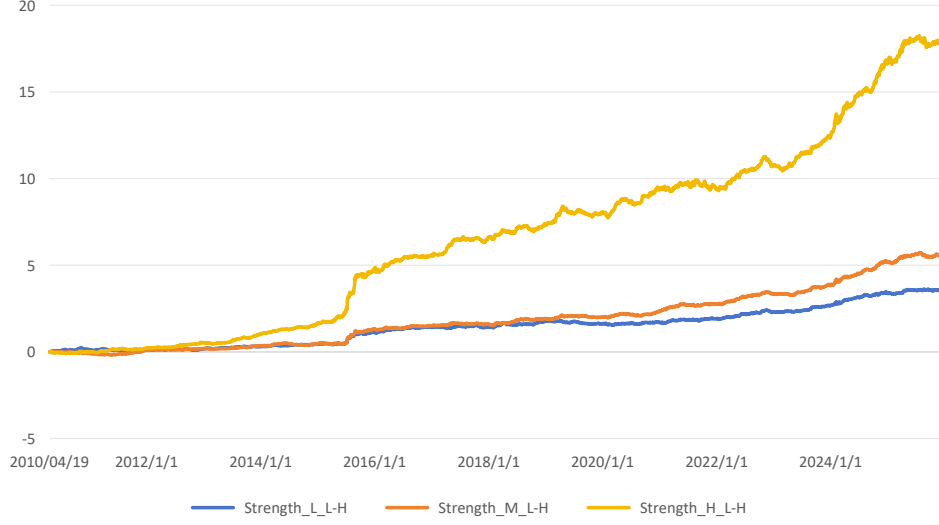


Figure 7: Cumulative Returns from the Double Sort on net_PCA and Sentiment Intensity

intensity increases, both the return magnitude and statistical significance of the Low–High portfolio rise. In the high-sentiment-intensity group, the annualized return reaches 22.02%, with a t -statistic of 7.80. These results indicate that the predictive relation between net_PCA and future returns becomes stronger when investor sentiment is more intense. Consistent with this evidence, the cumulative return curves show that the Low–High portfolio in the high-sentiment-intensity group experiences the most pronounced cumulative growth, providing further support for the amplifying role of sentiment intensity in the return effect associated with net_PCA.

4.2.3. Double-Sorting Tests by Sentiment Direction

After examining the role of sentiment intensity, we further investigate whether the return predictability of net_PCA varies with the direction of investor sentiment. We again employ a double-sorting procedure. Specifically, based on the sentiment direction measure *Sentiment*, stocks are first classified into pessimistic, neutral, and optimistic sentiment groups. Within each sentiment group, stocks are then further sorted into three portfolios according to net_PCA, denoted by PCA1, PCA2, and PCA3. This procedure yields a 3×3 set of double-sorted portfolios, from which we construct the net_PCA-

based Low–High long–short portfolio. The empirical results are reported below.

Table 12: Double-Sorting Results for net_PCA and Sentiment Direction

	PCA1	PCA2	PCA3	L-H
<i>SENT_L</i>	-0.016 (-0.52)	-0.011 (-0.36)	-0.055* (-1.85)	0.040*** (5.15)
<i>SENT_M</i>	0.058** (2.07)	0.040 (1.48)	0.011 (0.39)	0.047*** (5.02)
<i>SENT_H</i>	0.160*** (6.07)	0.099*** (4.08)	0.083*** (3.05)	0.076*** (7.36)

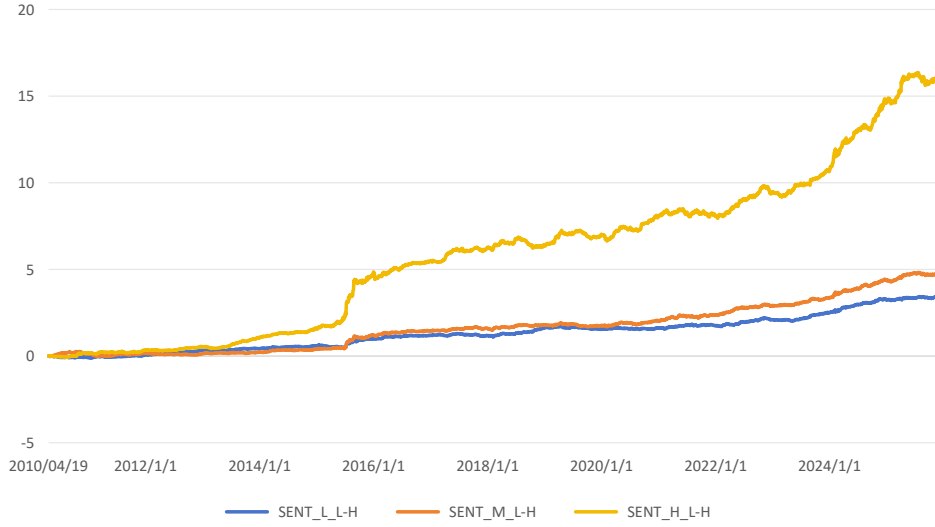


Figure 8: Cumulative Returns from the Double Sort on net_PCA and Sentiment Direction

Table 12 shows that the return of the Low–High portfolio increases steadily as investor sentiment shifts from pessimistic to optimistic, indicating that the predictive relation between net_PCA and future returns is strongest in the optimistic sentiment group. In this group, the annualized return reaches 21.19%, with a t -statistic of 7.36. The cumulative return curves exhibit the

same pattern: the cumulative long–short return in the optimistic sentiment group is substantially higher than those in the neutral and pessimistic groups.

These results suggest that the direction of investor sentiment is an important mechanism through which the pricing effect of net_PCA varies. When overall market sentiment is more optimistic, investors are more prone to overconfidence and return-chasing behavior, which can push stock prices further away from fundamental values. As a result, under more optimistic sentiment conditions, the sentiment-driven mispricing information embedded in net_PCA becomes more pronounced, subsequent price corrections are stronger, and the Low–High portfolio consequently earns larger returns.

4.3. Mechanism Analysis Based on Lottery Preference and Speculative Trading

4.3.1. Construction of the Lottery Preference Measure

We use the MAX measure to capture investors’ lottery preference. Behavioral finance theory suggests that some investors do not make investment decisions solely on the basis of conventional risk–return trade-offs, but instead favor stocks that offer a small probability of extremely high payoffs, reflecting lottery-like investment preferences. Such investors tend to be attracted to stocks that have recently experienced extreme positive returns and may be willing to pay a premium for the possibility of large gains. Accordingly, the historical incidence of extreme positive returns can serve as a proxy for investors’ preference for lottery-like stocks.

Following the literature on lottery preference, we construct the MAX measure as the maximum daily return over the previous 20 trading days. To avoid incorporating future information into the measure, MAX at trading day t is calculated using stock returns from the 20 trading days preceding t :

$$\text{MAX}_{i,t} = \max(r_{i,t-20}, r_{i,t-19}, \dots, r_{i,t-1}) \quad (13)$$

where $r_{i,t}$ denotes the daily return of stock i on trading day t , and $\text{MAX}_{i,t}$ denotes the lottery-preference measure for stock i at time t . A higher value of MAX indicates that the stock is more likely to attract investors seeking short-term extreme gains, implying stronger lottery preference and more pronounced speculative trading tendencies.

4.3.2. Double-Sorting Tests by Lottery Preference

We first sort stocks into low-, medium-, and high-lottery-preference groups based on the MAX measure. Within each MAX group, stocks are then further

sorted into three portfolios according to net_PCA. By comparing the returns of the Low–High portfolio across different levels of lottery preference, we assess whether lottery preference and speculative trading amplify the return predictability of net_PCA. The empirical results are reported below.

Table 13: Double-Sorting Results for net_PCA and Lottery Preference

	PCA1	PCA2	PCA3	L-H
<i>MAX_L</i>	0.067*** (2.69)	0.046* (1.95)	0.017 (0.66)	0.050*** (5.21)
<i>MAX_M</i>	0.078*** (2.74)	0.043 (1.50)	0.029 (0.98)	0.050*** (5.99)
<i>MAX_H</i>	0.058* (1.88)	0.025 (0.79)	-0.008 (-0.27)	0.067*** (6.53)

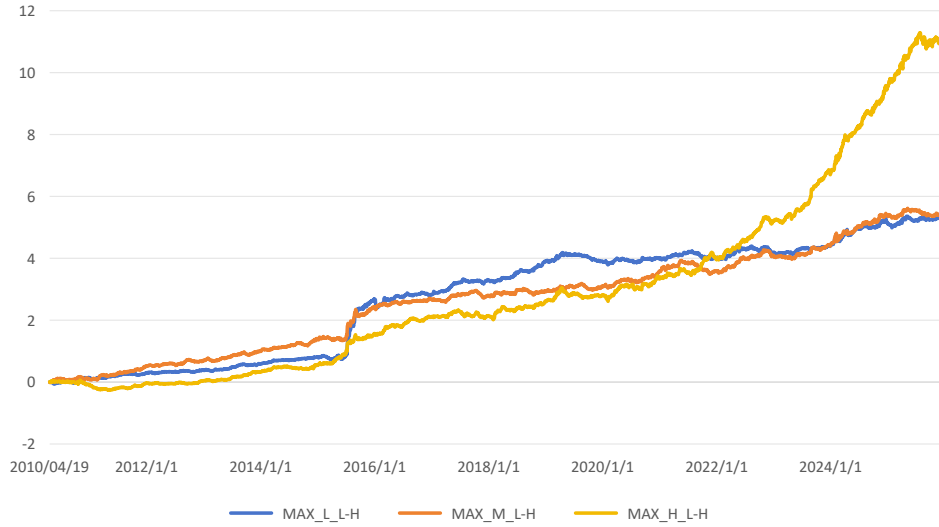


Figure 9: Cumulative Returns from the Double Sort on net_PCA and Lottery Preference

Table 13 shows that the Low–High portfolios constructed on the basis of net_PCA generate significantly positive returns across all MAX groups. This result indicates that net_PCA retains return predictability regardless of the

degree of lottery-like characteristics exhibited by the stocks. Importantly, both the return magnitude and statistical significance of the Low-High portfolio are strongest in the high-MAX group, indicating that the return predictability of net_PCA is more pronounced among stocks with stronger lottery-like characteristics.

The cumulative return curves provide further evidence consistent with this interpretation. The cumulative returns of the low-MAX and medium-MAX groups follow broadly similar paths, whereas the high-MAX group begins to rise more rapidly in the later part of the sample and ultimately ends well above the other two groups. This pattern suggests that high-MAX stocks are more likely to attract investors with a preference for extreme positive payoffs. When the optimistic expectations reflected in margin trading and short-selling activity coincide with lottery preference, stock prices may be more prone to deviate from fundamental values, leading to more pronounced return reversals as prices subsequently correct.

4.4. Discussion of Mechanism Analysis Results

Taken together, the mechanism tests suggest that the return predictability of net_PCA is not primarily attributable to simple risk compensation, but is instead more closely associated with short-term mispricing arising from investors' irrational trading behavior. On the one hand, the sentiment-based double-sorting results show that both the return magnitude and statistical significance of the net_PCA-based Low-High portfolio increase markedly with investor sentiment intensity, suggesting that stronger sentiment amplifies the mispricing information embedded in margin trading and short-selling activity. On the other hand, the MAX-based double-sorting results show that the predictive power of net_PCA is more pronounced among stocks with stronger lottery-like characteristics, indicating that speculative trading demand may further magnify price deviations and lead to subsequent return corrections.

A plausible economic interpretation is that margin trading and short-selling activity reflects not only investors' expectations about future stock price movements, but also non-fundamental information related to investor sentiment, risk preferences, and trading motives. When market sentiment is strong, whether strongly optimistic or strongly pessimistic, investors are more likely to engage in boundedly rational and herding behavior, making stock prices more prone to deviate from fundamental values. In this setting, the disagreement in investor expectations and sentiment preferences reflected

in margin trading and short-selling activity is more intensively captured by the `net_PCA` factor, thereby strengthening its predictive relation with future returns.

More specifically, when lottery preference is stronger in the market, investors are more likely to focus on stocks with extreme return characteristics and to pay a premium for the possibility of large gains. In this setting, the leverage embedded in margin financing may further amplify speculative demand, increasing the extent to which stock prices deviate from fundamental values. As a result, stocks with stronger lottery-like characteristics are more likely to accumulate price distortions generated by speculative trading and to experience more pronounced return reversals as market forces subsequently rebalance.

From the construction of `net_PCA`, the factor jointly extracts information from the margin-financing and short-selling sides and therefore captures the relative variation in bullish and bearish trading activity. When market sentiment is weak and investor behavior is relatively rational, margin trading and short selling are more likely to reflect normal responses to publicly available information, leaving less mispricing for `net_PCA` to capture. By contrast, when market sentiment is elevated and speculative trading is more active, investor irrationality and price deviations become more pronounced, causing the behavioral information embedded in `net_PCA` to become more salient and, in turn, strengthening its return predictability.

The same mechanism can also help explain the pronounced strengthening of `net_PCA` returns during the severe market turbulence around 2015. During periods of rapid market appreciation and subsequent correction, investor sentiment, leveraged trading, and speculative demand all tend to intensify, making stock prices more susceptible to deviations from fundamental values driven by irrational trading. As a result, the sentiment and speculative-trading information embedded in margin trading and short-selling activity is more strongly reflected in `net_PCA`, and subsequent price corrections translate this mispricing into larger Low–High portfolio returns.

5. Robustness Tests

5.1. *Portfolio Return Tests with Transaction Costs*

The preceding empirical analysis shows that `net_PCA` has significant predictive power for future returns. However, the portfolio-sorting strategy requires periodic rebalancing based on factor values, and changes in portfolio

composition and weights inevitably generate transaction costs. This concern is particularly relevant for daily and weekly rebalancing, where ignoring trading costs may overstate the practical implementability of the strategy. We therefore incorporate transaction costs into the original portfolio returns to examine whether the profitability of the margin-short composite factor survives after accounting for trading frictions.

Specifically, we first calculate portfolio turnover based on the constituent stocks and portfolio weights of the D1 and D10 portfolios in each period. For equal-weighted portfolios, each stock receives an equal weight within its corresponding portfolio. For value-weighted portfolios, each stock's weight is determined by its tradable market capitalization relative to the total tradable market capitalization of the portfolio. Based on changes in portfolio constituents and weights between adjacent periods, we then compute portfolio turnover as follows:

$$\text{Turnover}_{p,t} = 1 - \sum_i \min(\omega_{i,t}, \omega_{i,t-1}) \quad (14)$$

where $\omega_{i,t}$ and $\omega_{i,t-1}$ denote the weights of stock i in portfolio p at times t and $t - 1$, respectively. The turnover measure captures the proportion of the portfolio that must be traded when rebalancing from the previous period to the current period. A higher turnover rate indicates larger changes in portfolio constituents or weights and therefore a greater impact of transaction costs on portfolio returns.

After obtaining portfolio turnover, we calculate portfolio returns net of transaction costs. We consider three transaction-cost scenarios of 1, 2, and 5 basis points and estimate the rebalancing cost in each period as the transaction-cost rate multiplied by portfolio turnover. This cost is then deducted from the corresponding gross portfolio return. Rather than separately modeling commissions, regulatory trading fees, stamp duties, and other trading frictions, we approximate these costs using a single all-in transaction-cost rate.

At time t , we form portfolios based on net_PCA and evaluate their returns in period $t + 1$. Accordingly, the portfolio return net of transaction costs can be expressed as:

$$R_{p,t+1}^{\text{net}} = R_{p,t+1} - c \times \text{Turnover}_{p,t} \quad (15)$$

where $R_{p,t+1}$ denotes the gross return of portfolio p in period $t + 1$, c denotes the all-in transaction-cost rate, and $\text{Turnover}_{p,t}$ denotes the turnover

of portfolio p when it is formed or rebalanced at time t . For the High–Low and Low–High long–short portfolios, both the D1 and D10 legs are rebalanced. Transaction costs are therefore deducted based on the sum of the turnover rates of the two legs. The corresponding net return can be expressed as:

$$R_{LH,t+1}^{\text{net}} = R_{LH,t+1} - c \times (\text{Turnover}_{D1,t} + \text{Turnover}_{D10,t}) \quad (16)$$

Regarding the transaction-cost assumptions, 1 basis point represents a low-cost scenario, 2 basis points a moderate-cost scenario, and 5 basis points a high-cost or stress scenario. Table 14 reports the portfolio return results after accounting for transaction costs.

Table 14: Low-High Portfolio Returns after Accounting for Transaction Costs

	Frequency	Weighting Method	Cost Scenario	Turnover (%)	Mean Return (%)	Newey-West t	Annual Return (%)	Sharpe Ratio
44	Daily	Equal-Weighted	No Transaction Costs	155.14	0.11	10.493	28.80	2.846
			1bp		0.10	9.07	24.89	2.46
			2bp		0.08	7.646	20.98	2.073
			5bp		0.04	3.373	9.25	0.915
		Value-Weighted	No Transaction Costs	163.30	0.09	5.691	22.28	1.453
			1bp		0.07	4.639	18.16	1.185
			2bp		0.06	3.587	14.05	0.916
			5bp		0.01	0.434	1.70	0.111
	Weekly	Equal-Weighted	No Transaction Costs	158.55	0.23	4.014	11.95	1.034
			1bp		0.21	3.737	11.13	0.963
			2bp		0.20	3.46	10.30	0.891
			5bp		0.15	2.629	7.83	0.677
Equal-Weighted		No Transaction Costs	165.25	0.25	3.145	12.96	0.802	
		1bp		0.23	2.936	12.10	0.749	
		2bp		0.22	2.728	11.24	0.695	
		5bp		0.17	2.102	8.66	0.536	

As shown in Table 14, the returns of the Low–High portfolio decline after transaction costs are taken into account, but remain positive in most cases. This finding suggests that the profitability associated with net_PCA is not entirely driven by the assumption of zero transaction costs.

Turning to the daily results, the equal-weighted Low–High portfolio earns an average return of 0.11% before transaction costs. Under transaction-cost assumptions of 1, 2, and 5 basis points, the average return declines to 0.10%, 0.08%, and 0.04%, respectively, but remains positive and statistically significant in all three cases. For the value-weighted portfolio, returns remain significantly positive under the 1- and 2-basis-point scenarios, but under the relatively high 5-basis-point cost assumption, the average return falls to 0.01%, with a t -statistic of only 0.434. This result indicates that the daily value-weighted strategy is more sensitive to transaction costs. By comparison, the weekly Low–High portfolio is relatively more robust after accounting for trading costs. Under both equal weighting and value weighting, the weekly strategy continues to generate positive returns with statistical significance across all three transaction-cost scenarios.

Overall, transaction costs reduce the returns of the Low–High portfolio, with larger cost assumptions leading to more pronounced declines in profitability. This pattern is largely attributable to the relatively high portfolio turnover. The results suggest that the net_PCA-based strategy remains economically implementable under low and moderate transaction-cost scenarios, although practical applications would require careful control of rebalancing frequency and portfolio turnover.

5.2. *Orthogonalization Tests*

The preceding empirical results show that the composite factor constructed from margin trading and short-selling information significantly predicts future stock returns. However, firm characteristics such as size, short-term reversal, and liquidity may also affect cross-sectional stock returns. If the margin–short composite factor substantially overlaps with these conventional characteristics, the earlier findings may not fully reflect the incremental information contained in margin trading and short-selling activity itself. To further assess the independence of the factor’s return predictability, we orthogonalize net_PCA with respect to the relevant characteristic variables and reintroduce the orthogonalized factor into the Fama–MacBeth cross-sectional regressions.

Specifically, we implement two orthogonalization procedures. First, we regress net_PCA on firm size ($\ln_size$) and short-term reversal (REV)

and use the residual from this regression as the orthogonalized factor that purges the effects of size and reversal. We denote this residualized factor by orth_net_PCA1 , which is constructed as follows:

$$\text{net_PCA}_{i,t} = \alpha_t + \beta_{1,t}\text{ln_size}_{i,t} + \beta_{2,t}\text{REV}_{i,t} + \varepsilon_{i,t} \quad (17)$$

$$\text{orth_net_PCA1}_{i,t} = \varepsilon_{i,t} \quad (18)$$

Second, we regress net_PCA on turnover and illiquidity and use the residual as the orthogonalized factor orth_net_PCA2 , which purges liquidity-related effects. We then re-estimate the Fama–MacBeth regressions using orth_net_PCA1 and orth_net_PCA2 as the key explanatory variables.

Table 15: Fama–MacBeth Regression Results after Removing Size and Reversal Effects

Variable	Model(1)	Model(2)	Model(3)
<i>Alpha</i>	0.04 (1.427)	0.080*** (3.001)	0.043* (1.705)
<i>orth_net_PCA1</i>	-0.015*** (-6.613)	-0.010*** (-4.539)	-0.007*** (-3.451)
<i>Turnover</i>		-2.159*** (-5.632)	-2.069*** (-5.246)
<i>ln_Amihud</i>			275.457*** (6.236)

The results show that the orthogonalized margin–short composite factor, orth_net_PCA1 , remains significantly negative across all three model specifications. In Model (1), the coefficient on orth_net_PCA1 is -0.015%, with a t -statistic of -6.613, indicating that the factor retains significant return predictability after removing the effects of firm size and short-term reversal. After further controlling for turnover and illiquidity, the coefficient remains significant at the 1% level. These findings suggest that the return predictability of the margin–short composite factor is not driven by size or short-term reversal effects. Moreover, the persistence of the orthogonalized factor after additional liquidity controls indicates that net_PCA contains incremental pricing information beyond conventional cross-sectional stock characteristics.

Table 16: Fama–MacBeth Regression Results after Removing Liquidity Effects

Variable	Model(1)	Model(2)	Model(3)
<i>Alpha</i>	0.040	0.142	0.087
	1.422	1.187	0.857
<i>orth_net_PCA2</i>	-0.015***	-0.014***	-0.004**
	(-6.435)	(-6.816)	(-2.015)
<i>ln_size</i>		-0.007	-0.005
		(-1.042)	(-0.877)
<i>REV</i>			3.1705***
			(10.731)

The results further show that *orth_net_PCA2* is significantly negative across all three model specifications. This finding indicates that even after removing the liquidity effects captured by turnover and illiquidity, the margin–short composite factor continues to exhibit significant return predictability. Therefore, the pricing effect of *net_PCA* is not merely driven by cross-sectional differences in stock liquidity, but instead contains incremental information that is distinct from conventional characteristics such as liquidity, firm size, and short-term reversal.

5.3. Robustness Tests across Market States

To further examine whether the return predictability of the margin–short composite factor depends on particular market conditions, we partition the sample according to the overall market state and re-run the portfolio sorting tests separately for bull and bear markets. Specifically, we use the CSI 300 Index as a proxy for market conditions. We first compute daily market returns from the closing level of the CSI 300 Index and then calculate the cumulative return over the previous 252 trading days. To avoid look-ahead bias, the market state on trading day t is determined using the cumulative return over the 252 trading days ending at $t - 1$. A trading day is classified as a bull market if this cumulative return is greater than zero and as a bear market if it is below zero.

Table 17 reports the decile-sorting results across different market states. During bear markets, the Low–High portfolio earns an annualized return of 30.860% with a Sharpe ratio of 3.420. During bull markets, the corresponding

Table 17: Equal-Weighted Portfolio Return Tests across Bull and Bear Markets

		Mean Daily Return (%)	Newey- West t	Annual Return (%)	Annual Volatil- ity(%)	<i>Sharpe Ratio</i>
bear	<i>D1</i>	0.080	2.0467	20.175	24.989	0.807
	<i>D10</i>	-0.042	-1.081	-10.685	26.308	-0.406
	<i>High-Low</i>	-0.122	-10.220	-30.86	9.022	-3.420
	<i>Low-High</i>	0.122	10.220	30.86	9.022	3.420
bull	<i>D1</i>	0.111	2.532	27.906	27.307	1.0219
	<i>D10</i>	0.005	0.108	1.217	29.179	0.0417
	<i>High-Low</i>	-0.106	-5.795	-26.688	11.137	-2.396
	<i>Low-High</i>	0.106	5.795	26.688	11.137	2.396

annualized return is 26.688%, with a Sharpe ratio of 2.396. These results show that the Low–High portfolio generates significantly positive returns in both bull and bear markets.

These results indicate that the return predictability of the margin–short composite factor does not depend on a particular market regime. During both rising and falling market conditions, the low-net_PCA portfolio significantly outperforms the high-net_PCA portfolio, suggesting that the return-predictive effect documented above remains robust across different market states.

6. Conclusion

Using data on stocks eligible for margin trading and short selling in China’s Shanghai and Shenzhen A-share markets from 2010 to 2025, this study examines whether margin trading and short-selling activity contains asset pricing information that has not yet been fully incorporated into stock prices. In contrast to studies that rely on individual measures such as margin balances, short-selling balances, or trading volume, we characterize margin-financing and short-selling activity along three dimensions—changes in balances, relative trading intensity, and sentiment momentum—and employ principal component analysis to extract the common information on each side. We then construct a margin–short composite factor that captures the relative

variation in margin-financing and short-selling activity. Using portfolio sorting tests, risk-adjusted return analyses, holding-period tests, and Fama–MacBeth cross-sectional regressions, we systematically evaluate the factor’s return predictability and investigate its underlying economic mechanisms.

The results show that the margin–short composite factor significantly and negatively predicts future stock returns. Stocks with lower factor values earn higher subsequent returns, whereas stocks with higher factor values earn lower future returns. In both the daily and weekly samples, a strategy that goes long the low-factor portfolio and short the high-factor portfolio generates significantly positive returns. The equal-weighted and value-weighted results are directionally consistent, indicating that the effect is not entirely driven by small-cap stocks. After adjusting for risk using the CAPM, the Fama–French three-factor model, and the Fama–French five-factor model, the long–short portfolio continues to earn significantly positive alphas. Moreover, after controlling for firm size, short-term reversal, turnover, and illiquidity, the Fama–MacBeth coefficient on the composite factor remains significantly negative. These findings indicate that the margin–short composite factor contains incremental return-predictive information beyond conventional risk factors and commonly used firm characteristics.

The return predictability of the margin–short composite factor exhibits a clear horizon-dependent pattern. The factor performs strongly at the daily and weekly frequencies, but its predictive power weakens substantially at the monthly frequency. As the holding horizon extends from 1 to 30 trading days, both the annualized return and the Sharpe ratio of the long–short portfolio decline gradually. These results suggest that the return differences associated with margin trading and short-selling information are more consistent with short-term price deviations and subsequent corrections than with persistent compensation for bearing long-run systematic risk. Because a higher composite-factor value indicates stronger margin-financing activity relative to short-selling activity, stocks with high factor values should earn higher subsequent returns if margin financing primarily reflects favorable private information that has not yet been incorporated into prices. Instead, we find that high-factor stocks subsequently earn lower returns, suggesting that the optimistic trading demand reflected in relatively stronger margin-financing activity may already have pushed current prices upward and is followed by subsequent return reversals.

The mechanism tests further show that this predictive relation is consistent with mispricing arising from investor sentiment and speculative trading.

Sentiment measures constructed from stock forum comments indicate that the long–short return of the margin–short composite factor increases markedly with investor sentiment intensity, and that the factor’s return predictability also strengthens as sentiment shifts from pessimistic to optimistic. At the same time, the long–short portfolio earns relatively higher returns among stocks with high MAX characteristics, suggesting that investors’ preference for extreme positive payoffs may amplify the optimistic demand facilitated by margin leverage and make stock prices more prone to deviate from fundamental values. Importantly, these findings should be interpreted as evidence consistent with a behavioral mispricing explanation: investor sentiment and lottery preference appear to strengthen the return effect of the factor, but the results do not constitute a complete identification of any single causal channel.

The main findings remain robust across a range of additional tests. After accounting for different levels of transaction costs, most daily and weekly long–short portfolios continue to generate positive returns, with the weekly strategy showing relatively lower sensitivity to trading costs. When the composite factor is orthogonalized with respect to firm size and short-term reversal, as well as turnover and illiquidity, the orthogonalized factors continue to exhibit significantly negative return predictability. In addition, the low-factor portfolio significantly outperforms the high-factor portfolio in both bull and bear markets, indicating that the effect is not confined to a particular market regime. At the same time, the relatively high portfolio turnover implies that the statistical returns associated with the margin–short composite factor should not be interpreted as equivalent to the net returns attainable by investors in practice. Practical implementation would also need to account for rebalancing frequency, market impact costs, and the availability of securities for short selling.

Overall, our findings show that margin trading and short-selling data are not merely outcomes of market trading activity, but also contain cross-sectional information that can be used to identify short-term mispricing. By integrating multidimensional trading characteristics from both the margin-financing and short-selling sides, this study provides new empirical evidence extending the literature on margin trading and short selling from institutional effects to asset pricing and return predictability. The results also suggest that high-frequency trading-behavior data can serve as an important complement to traditional accounting information and price–volume data. From a practical perspective, the margin–short composite factor may help investors identify excessive optimism and potential subsequent price corrections. However, its

usefulness is concentrated primarily at shorter horizons and is constrained by real-world frictions such as transaction costs, portfolio turnover, and short-sale constraints.

References

- Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5, 31–56.
- Antweiler, W., Frank, M.Z., 2004. Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance* 59, 1259–1294.
- Asquith, P., Pathak, P.A., Ritter, J.R., 2005. Short interest, institutional ownership, and stock returns. *Journal of Financial Economics* 78, 243–276.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. *The Journal of Finance* 61, 1645–1680.
- Bali, T.G., Cakici, N., Whitelaw, R.F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99, 427–446.
- Barber, B.M., Odean, T., Zhu, N., 2008. Do retail trades move markets? *The Review of Financial Studies* 22, 151–186.
- Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. *Journal of Financial Economics* 49, 307–343.
- Boehme, R.D., Danielsen, B.R., Sorescu, S.M., 2006. Short-sale constraints, differences of opinion, and overvaluation. *Journal of Financial and Quantitative Analysis* 41, 455–487.
- Boehmer, E., Wu, J., 2013. Short selling and the price discovery process. *The Review of Financial Studies* 26, 287–322.
- Chang, E.C., Luo, Y., Ren, J., 2014. Short-selling, margin-trading, and price efficiency: Evidence from the Chinese market. *Journal of Banking & Finance* 48, 411–424.
- Chen, J., Hong, H., Stein, J.C., 2002. Breadth of ownership and stock returns. *Journal of Financial Economics* 66, 171–205.

- Chu, J., Fang, J., 2016. Chinese-style margin trading and securities lending arrangements and the worsening of stock price crash risk. *Economic Research Journal* 51, 143–158. In Chinese; title translated from the original.
- Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under-and overreactions. *The Journal of Finance* 53, 1839–1885.
- Datar, V.T., Naik, N.Y., Radcliffe, R., 1998. Liquidity and stock returns: An alternative test. *Journal of Financial Markets* 1, 203–219.
- De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Noise trader risk in financial markets. *Journal of Political Economy* 98, 703–738.
- Diamond, D.W., Verrecchia, R.E., 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics* 18, 277–311.
- Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. *The Journal of Finance* 47, 427–465.
- Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.
- Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116, 1–22.
- Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81, 607–636.
- Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.
- Goyal, A., Reed, A.V., Smajlbegovic, E., Soebhag, A., 2025. Stealthy shorts: Informed liquidity supply. *Journal of Financial Economics* 172, 104155.
- Greenwood, R., Shleifer, A., 2014. Expectations of returns and expected returns. *The Review of Financial Studies* 27, 714–746.

- Gu, Z., Hao, X., Zhang, Y., 2011. Short-sale constraints, investor behavior, and pricing bubbles in the A-share market. *Journal of Financial Research* , 129–148. In Chinese; title translated from the original.
- He, C., Chen, R., Xue, B., et al., 2021. Investor sentiment, limits to arbitrage, and stock price anomalies. *Economic Research Journal* 56, 58–73. In Chinese; title translated from the original.
- Kelly, B.T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics* 134, 501–524.
- Kumar, A., 2009. Who gambles in the stock market? *The Journal of Finance* 64, 1889–1933.
- Kyle, A.S., 1985. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society* , 1315–1335.
- Li, K., Xu, L., Zhu, W., 2014. Short-sale constraints and stock mispricing: Evidence from the margin trading and securities lending system. *Economic Research Journal* 49, 165–178. In Chinese; title translated from the original.
- Li, Z., Chen, C., Lin, B., 2015. Does short selling improve pricing efficiency in the Chinese stock market? Evidence from a natural experiment. *Economic Research Journal* 50, 165–177. In Chinese; title translated from the original.
- Miller, E.M., 1977. Risk, uncertainty, and divergence of opinion. *The Journal of Finance* 32, 1151–1168.
- Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Seguin, P.J., 1990. Stock volatility and margin trading. *Journal of Monetary Economics* 26, 101–121.
- Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. *The Journal of Finance* 52, 35–55.
- Stambaugh, R.F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104, 288–302.

- Wan, X., 2024. Margin-buying, short-selling, and stock valuation: Why is the effect reversed over time in China? *Journal of Empirical Finance* 76, 101476.
- Xiao, H., Kong, A., 2014. The mechanism through which margin trading and securities lending affect idiosyncratic stock volatility: A difference-in-differences test. *Management World* , 30–43, 187–188In Chinese; title translated from the original.
- Xu, H., Chen, X., 2012. Does the introduction of margin trading and securities lending improve the pricing efficiency of eligible stocks in China? A difference-in-differences study. *Management World* , 52–61In Chinese; title translated from the original.