Analysis of the Pricing Capability of the Expanded Fama-French Five-Factor Model Based on Investor Sentiment

Yuliang Li

Stirling College, Chengdu University, Chengdu, 610106, China

202216801525@cdu.edu.cn

Abstract. The Chinese A-share market has a distinctive market ecosystem, with retail investors accounting for over 99%. Their unique trading behaviors significantly influence market liquidity and the formation mechanism of prices. Meanwhile, traditional asset pricing models have limitations when applied in emerging markets, and they are unable to explain abnormal returns caused by investors' irrational behaviors. This paper constructs an investor sentiment factor based on market trading behaviors, technical indicators, and alternative data, and extends the Fama-French five-factor model. Three methods, namely time series regression, GRS test, and Fama-MacBeth cross-sectional regression, are used to test the model. The results show that the investor sentiment factor has stable pricing ability in the emerging market environment with a high proportion of retail investors and significant sentiment fluctuations. It effectively explains abnormal returns that traditional models fail to capture. Additionally, institutional investors can use the sentiment factor to optimize investment portfolios and risk management, while individual investors can reduce irrational investments based on it. Regulatory authorities can also build a market monitoring system to prevent systemic risks.

Keywords: A-share Market, Investor Sentiment, Fama-French Five-Factor Model, Asset Pricing, Behavioral Finance.

1. Introduction

The Chinese A-share market, currently the second-largest stock market globally, has developed over the past three decades and has formed a unique market ecosystem. As of 2024, there are 5,392 listed companies in the domestic stock market, with a total market value of 85.98 trillion yuan. In 2024, the cumulative number of new A-share stock accounts opened reached 24.9989 million, an increase of 16.6% compared to 21.4436 million in 2023. Among them, individual investors account for over 99%, and their market sentiment and irrational behaviors significantly influence market liquidity and volatility. The emotions and non-rational actions of individual investors have become an indispensable factor in price determination.

Under the interaction of the trading mechanism and investor structure in the Chinese market, complex and even counterproductive results may arise. Some listed companies use stock splits as a favorable event to manipulate the market. That is, insiders raise the stock price by releasing split announcements, attracting retail investors to follow suit and buy, and then insiders reduce their holdings at high prices through block trades or pledge loans, etc. (Titman, et. al, 2022). This behavior intensifies market volatility and irrational trading, making it increasingly difficult to accurately select stocks for excess returns through traditional pricing models, and posing higher demands on investors' trading.

Asset pricing theory is the cornerstone of multi-factor quantitative investment. The mean-variance framework was the first to mathematize investment decision-making problems, defining expected returns as the mean and investment risk as variance (Markowitz, 1952; Carhart, 1997). Based on this theory, Sharpe proposed the capital asset pricing framework, which established a linear relationship between the expected return of an asset and market risk, initiating the era of single-factor asset pricing. However, due to its limited explanatory power for market anomalies, the academic community has continuously sought risk factors that can more comprehensively explain excess returns. Therefore, Fama and French proposed a three-factor pricing model, which, on the basis of the market factor, added the size factor and value factor (Fama and French, 1993). To further explain the impact of profitability and investment style on returns, Fama and French incorporated the profitability factor and investment factor into the three-factor pricing model, forming the widely used five-factor pricing model (Fama and French, 2015).

However, most of the classic multi-factor asset pricing models originated from mature markets, and their application in emerging markets has significant limitations. Relevant research has found that the traditional pricing system performs well in developed markets but poorly in emerging markets (Kostin, et. al, 2022). Therefore, exploring factors that can capture the specific risks of the local market has become the key to improving the pricing framework's performance. Against this backdrop, the importance of the investor sentiment factor is increasingly prominent. Relevant research results indicate that the difference in investor sentiment is the key point to explain the abnormal returns of A-shares, and it is found that the amplification effect of idiosyncratic risk on premium will significantly increase when sentiment is high (Li and Zhang, 2021). This indicates that sentiment is a quantifiable and significantly explanatory pricing factor. In the A-share market, quantifying sentiment using various technical indicators and alternative data and incorporating it into pricing models is of vital importance. Although previous studies have attempted to construct investor sentiment factors, there has been insufficient exploration in their integration with traditional models, which is precisely the main problem this paper aims to solve.

The main content of this article consists of three parts. Firstly, based on multiple dimensions such as technical indicators, market sentiment, and capital flow, the principal component analysis method is used to construct investor sentiment factors. Descriptive statistics, stationarity tests, correlation analysis, and multicollinearity diagnosis are employed to ensure their validity. Subsequently, the sentiment factors are integrated into the traditional five-factor pricing framework, and the load coefficients and intercept terms of each factor are estimated using the time-ordered regression method. Finally, by systematically comparing the differences between the two models in terms of Wald tests and GRS tests, the incremental contribution of the sentiment factors is evaluated.

2. Research Methods

2.1. Method Introduction

Fama and French proved that market risk compensation, size effect, and value effect are the key factors influencing the cross-sectional returns of stocks. After incorporating size and value factors, the explanatory power for the differences in asset returns was enhanced (Fama and French, 1993). However, subsequent studies pointed out their limitations, particularly the difficulty in explaining the changes in excess returns of stocks with small size and low valuation (Yao, 2023). Therefore, the two constructed a five-factor model, significantly improving the explanatory power (Fama and French, 2015). This system is based on the MM theorem and the cash flow discount pricing logic, and holds that companies with high profits and low investment will receive higher expected returns.

However, the market efficiency hypothesis and the rational person assumption on which traditional models rely are facing real-world challenges. Empirical evidence shows that they cannot explain the abnormal returns resulting from investors' irrational behaviors, and these irrational factors can contribute to the formation of factor premiums (Habibah, et. al, 2021). Relevant studies on emotional factors indicate that emotional changes are more pronounced in stocks with opaque pricing, poor liquidity, and limited arbitrage opportunities (Wurgler and Baker, 2006). These studies suggest that emotions should be included as independent risk factors in the model.

In the Chinese A-share market, the proportion of individual investors is relatively high, and the impact of emotional fluctuations on trading decisions is more pronounced. In this environment, the traditional pricing framework shows certain limitations. Gao et al.'s research based on web text and trading data found that sentiment indicators are significantly negatively correlated with stock returns (Gao, et. al, 2024), confirming that sentiment is a systematic factor in A-share asset pricing. From the perspective of behavioral finance, investor cognitive biases and sentiment fluctuations lead to price deviations from fundamentals, constituting a priceable risk. Moreover, sentiment factors can explain anomalies such as the momentum effect, and their inclusion in models can enhance the explanatory power of abnormal A-share returns (Zhou, et. al, 2025).

Based on this, in this paper, on the basis of the traditional framework, this study introduces the investor sentiment factor to construct an expanded five-factor pricing system. The formula is as follows.

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i MKT_t + \beta_i SMB_t + \beta_i HML_t + \beta_i RMW_t + \beta_i CMA_t + \beta_i SENT_t + \epsilon_{i,t} \quad (1)$$

Among them, the emotional factor SENT is used to measure the overall emotional level of investors, capturing the systematic risks caused by the collective emotional fluctuations of investors. The load coefficient βi reflects the sensitivity of asset i to the risks brought about by emotional fluctuations, while the risk premium of the emotional factor represents the compensation income given by the market for bearing such risks.

2.2. Factors and Model Testing Methods

This paper employs a variety of statistical testing methods to ensure the validity of the data and the robustness of the model. Firstly, the information coefficient (IC) and rank information coefficient (RankIC) are used to evaluate the predictive effect of the factors on future returns. Compared to IC, RankIC is less affected by extreme values and thus is more robust in empirical studies. Secondly, VIF is used to detect multicollinearity. If the VIF value is high, it indicates a strong linear correlation among the factors, and they need to be dealt with. At the same time, the ADF test is used to confirm the stability of the factor sequence, avoiding the problem of false regression.

This paper employs the ordinary least squares method to conduct time-ordered regression on two asset pricing models, in order to estimate the factor loadings and intercept terms of each investment portfolio. To evaluate the incremental effect of the sentiment factor and the superiority of the extended model, this paper uses two methods for comparison. The GRS test is used to assess the validity of the pricing system. If the test fails, it indicates that there are systematic errors in this framework, and its explanatory power for excess returns is limited. If the test passes, it suggests that this framework can better explain the changes in returns. The Fama-MacBeth regression is conducted in two steps: first, a time-ordered regression is performed between stock excess returns and factors to obtain factor exposure, and then a cross-sectional regression is conducted to obtain the risk premium, and the significance of the factors is determined by calculating the average value and standard error of the results. The combination of the two can be used to judge the incremental contribution of the sentiment factor and the model performance, providing empirical evidence.

2.3. Data Sources and Processing

The data in this article is sourced from the iFind database, Baidu Search Index and the China Currency Network. The sample covers the daily trading data of the CSI 300 stocks from September 1, 2021 to September 1, 2025, as well as the search index, the CSI 300 ETF data and the one-year treasury bond yield. The final expanded five-factor dataset consists of 7 columns and 254,070 rows.

Null value handling removes the unlisted and undisclosed quarterly financial data, leaving 282 stocks. When the Baidu search index is missing, the stock code index is used as a substitute, or the relevant index is weighted and integrated to complete the data. Outlier handling performs 1% and 99% quantile Winsorization on non-yield continuous variables and corrects the values of the rise and fall stop of the yield. Additionally, perform one-hot encoding on the whether rise or fall stop limit column that contains four states of limit-up, limit-down, delist and no limit-up/limit-down.

3. Empirical Results

3.1. Validity Analysis of Candidate Factors and Construction of Investor Sentiment Factors

This paper conducts descriptive statistics on 19 candidate sentiment factors. The fluctuations and distributions of different factors are significantly different. For instance, the standard deviation of the momentum indicator is high, the high percentile of the turnover rate indicates trading frenzy, and the valuation indicators are right-skewed. The ADF test shows that most factors are stationary at the 1% level, while a few individual stocks are non-stationary. The Pearson analysis reveals that some factors are highly correlated, such as the turnover rate and the daily trading volume relative to market capitalization ($|r| > 0.8$), and the valuation and trading-related factors have low correlations. The VIF test shows that some factors have severe multicollinearity (VIF > 10), and the removal of redundant factors lays the foundation for subsequent principal component analysis.

This paper follows the method of Baker and Wurgler and uses principal component analysis (PCA) to extract the emotional dimension from five factors such as momentum indicators and abnormal turnover rate. PCA shows that the eigenvalue of PC1 is 1.70, the variance contribution rate is 34.04%, and the cumulative proportion of the first three items reaches 74.50%, as shown in Figure 1 and 2. PC1 has high loadings on abnormal turnover rate and amplitude, representing the market's speculative activity and emotional fluctuations. According to Gniazdowski's viewpoint, this paper retains PC1 as an emotional factor (SENT) (Gniazdowski, 2021). After standardization processing, the standard deviation of this indicator is 1 and the mean is 0. Positive values indicate an optimistic emotional bias, while negative values indicate a pessimistic emotional bias, thereby ensuring the objectivity and robustness of the measurement.

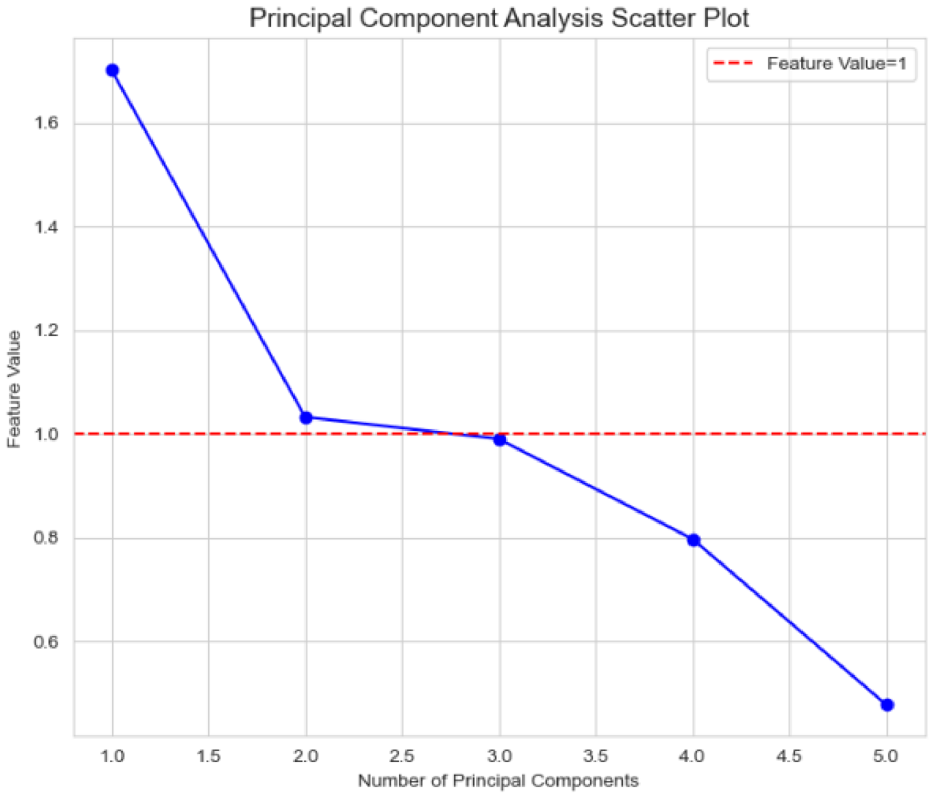

Figure 1. Principal component factor loading matrix.

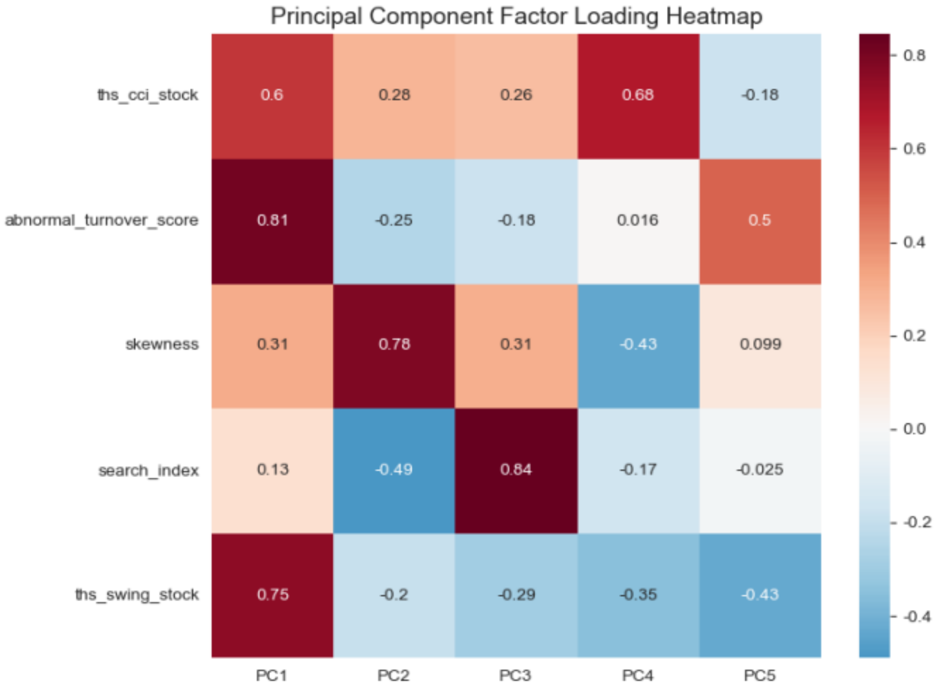

Figure 2. Principal component factor gravel plot.

3.2. Extended Five-Factor Validity Analysis

Based on the traditional five-factor pricing framework, this paper introduces the investor sentiment factor (SENT) and constructs an expanded six-factor asset pricing system. As shown in Table 1, the means of the traditional five factors (MKT, SMB, HML, RMW, CMA) are all close to 0, and the volatility levels are in line with theoretical expectations. Among them, the mean of the investment factor (CMA) is slightly positive, indicating that the conservative investment portfolio achieved a slight excess return during the sample period.

In contrast, SENT exhibits significantly different statistical characteristics. Its mean is negative, indicating that the overall market sentiment is pessimistic. The standard deviation is large, suggesting extremely high volatility of sentiment. The distribution is significantly skewed and there are extreme values, which verifies the intense fluctuation characteristic of sentiment in the market. Overall, the inclusion of SENT introduces asymmetric characteristics of behavioral finance to the model.

Table 1. Overview of expanded Five-Factor and Excess Return data.

Factor Name	Mean	Std	Min	25%	Median	75%	Max
MKT	-0.0074	0.0101	-0.0337	-0.0136	-0.0076	-0.0017	0.0255
SMB	-0.0004	0.0048	-0.0130	-0.0035	-0.0004	0.0027	0.0122
HML	-0.0004	0.0082	-0.0211	-0.0057	-0.0000	0.0052	0.0199
RMW	-0.0003	0.0052	-0.0137	-0.0037	-0.0002	0.0030	0.0125
CMA	0.0002	0.0055	-0.0146	-0.0032	0.0001	0.0037	0.0150
SENT	-0.0144	0.1274	-2.0413	-0.9352	-0.2556	0.6381	4.1534
excess_return	-0.0183	0.0242	-0.2230	-0.0305	-0.0193	-0.0079	0.1895

The results of the Fisher combined ADF test in this paper indicate that the ADF statistics of all factors are negative, and the p-values are much smaller than 0.05, with the proportion of stationary sequences reaching 100%. This means that all factors are stationary sequences and can be directly used for time-ordered regression and asset pricing analysis, without the need for differencing.

As shown in Figure 3, there are significant structural relationships among the traditional factors. Among them, HML is highly positively correlated with CMA, while RMW is negatively correlated with both, which conforms to the theoretical opposition between value and quality. MKT has the highest correlation with the excess return rate, verifying the core role of market risk.

The SENT factor exhibits unique characteristics. It has a moderate correlation with MKT ($r = 0.15$) and a higher correlation with excess returns ($r = 0.34$), indicating a significant connection between emotions and returns. At the same time, it has a weak correlation with other style factors, suggesting that it provides an independent source of information.

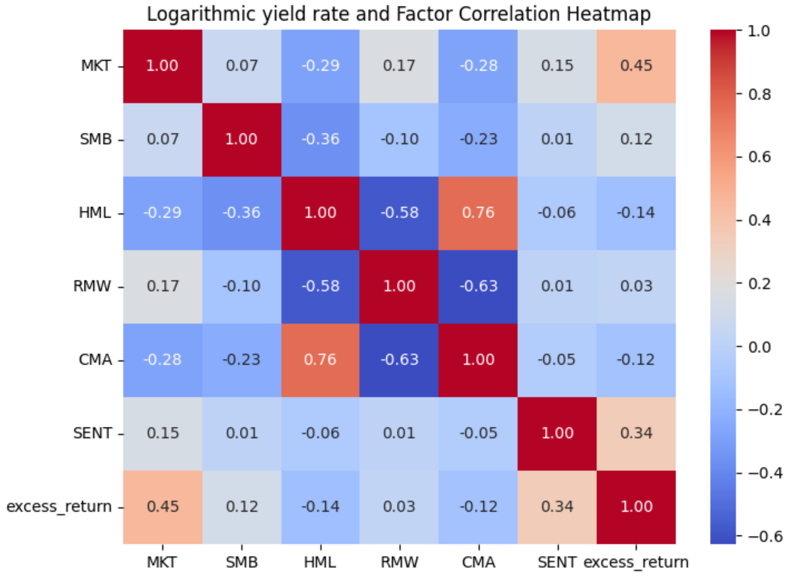

Figure 3. Correlation matrix of Excess Returns and each factor in the Extended Five-Factor dataset.

In the Pearson and Spearman correlation analyses, the MKT factor exhibited a strong linear correlation, while the SENT factor might have non-linear characteristics. Its Spearman coefficient was relatively low, indicating that its monotonicity was weaker than the linear relationship. The results of the two correlation coefficients for the remaining factors were basically consistent, suggesting that their relationship with the returns mainly presented a linear feature.

To further verify the stability of the model, a VIF test was conducted in this study. The results are shown in Table 2. It can be seen that the VIF values of all factors are below 5, indicating that there are no obvious collinearity problems in this framework. Among them, the test result of the emotional factor is approximately 1.02, indicating that its linear dependence with other risk factors is extremely low. This further verifies the independence and rationality of the SENT factor.

Table 2. Expanded Five-Factor IC analysis, IC ranking, multicollinearity analysis, and average correlation.

Factor Name	IC	Rank IC	VIF	Average Correlation
MKT	0.4545	0.4694	1.0827	0.1899
SENT	0.3396	0.2217	1.0194	0.0542
SMB	0.1171	0.0768	1.3802	0.1550
RMW	0.0334	0.0281	2.0613	0.2968
CMA	-0.1165	-0.1016	2.7404	0.3890
HML	-0.1354	-0.1129	2.8046	0.4091

3.3. Expansion of the Five-Factor Model Test

3.3.1 Time Series Regression Test

In the univariate panel regression, all traditional factors as well as the newly added investor sentiment factor (SENT) were included in the regression equation. The results of the time series regression showed that the SENT factor was statistically significant ($p < 0.05$) for the abnormal returns of 281 stocks in the sample. The distribution of SENT β is as shown in Figure 4, with the data concentrated around positive values, and the mean is 0.0045. The t-value distribution is as shown in Figure 5, and this value significantly exceeded the ±2 threshold, further supporting the conclusion that the sentiment factor serves as a systematic explanatory variable. The results indicated that the sentiment factor exhibited significant risk pricing characteristics in both cross-sectional regression and time dimension tests.

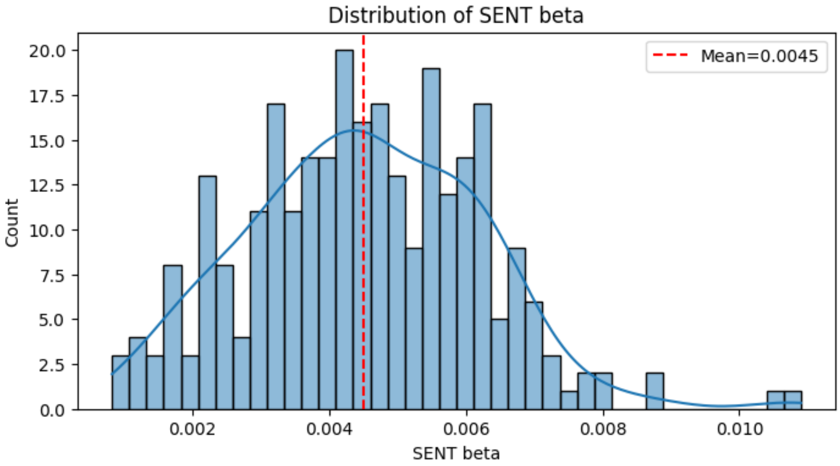

Figure 4. Histogram of factor loadings for investor sentiment factors

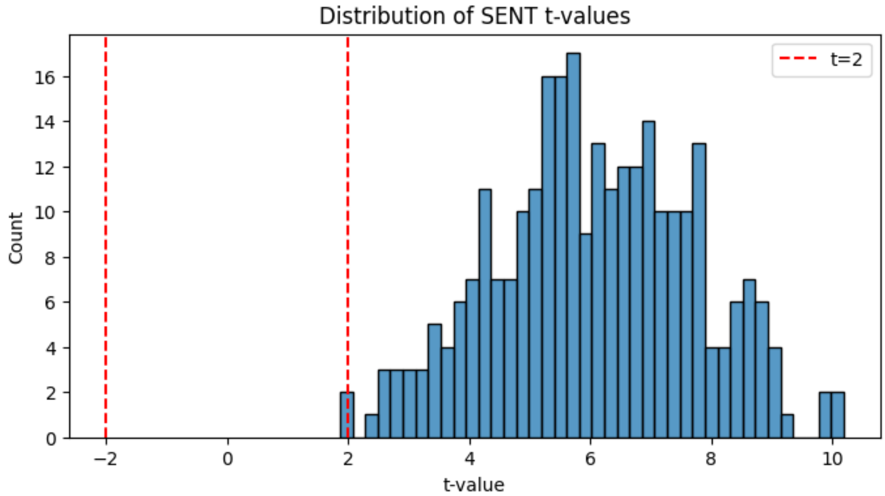

Figure 5. Histogram of t-values distribution

3.3.2 GRS Test

The GRS test results show that the statistical value of the traditional model is 24.76, and the p-value is close to 0 (1.11e-16), indicating that the overall alpha is not simultaneously zero, and there is a significant systematic pricing error. The GRS statistical value of the expanded five-factor model drops to 19.87, and the p-value is still close to 0, suggesting that the overall alpha is not completely zero, but compared to the traditional model, the pricing error has significantly converged, and the overall alpha has tended to decrease. However, as shown in Figure 6 and 7, although the alpha of the expanded five-factor model is not zero overall, compared to the five-factor model, the alpha has significantly converged, and the pricing error has significantly decreased. This indicates that the SENT factor has improved the consistency of asset pricing across cross-sections and has made an important contribution to the model improvement, but the model still cannot fully explain the abnormal returns of all assets.

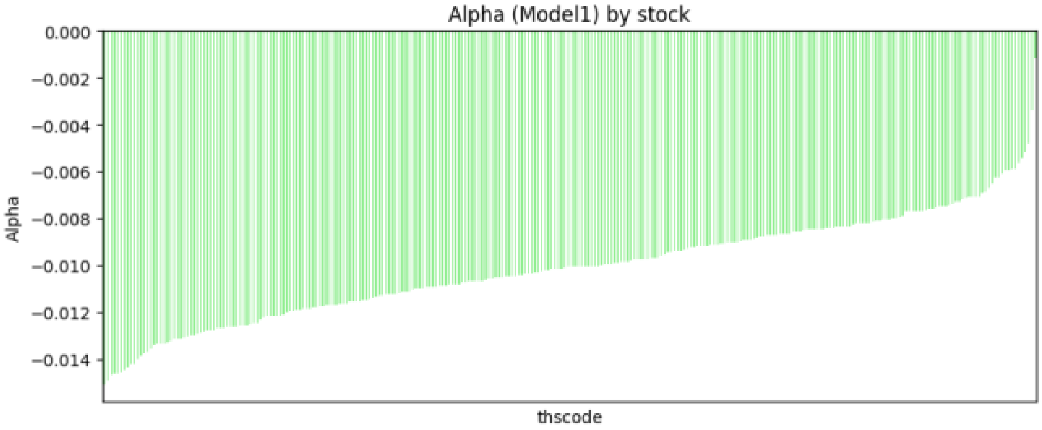

Figure 6. Histogram of intercept values of the five-factor model

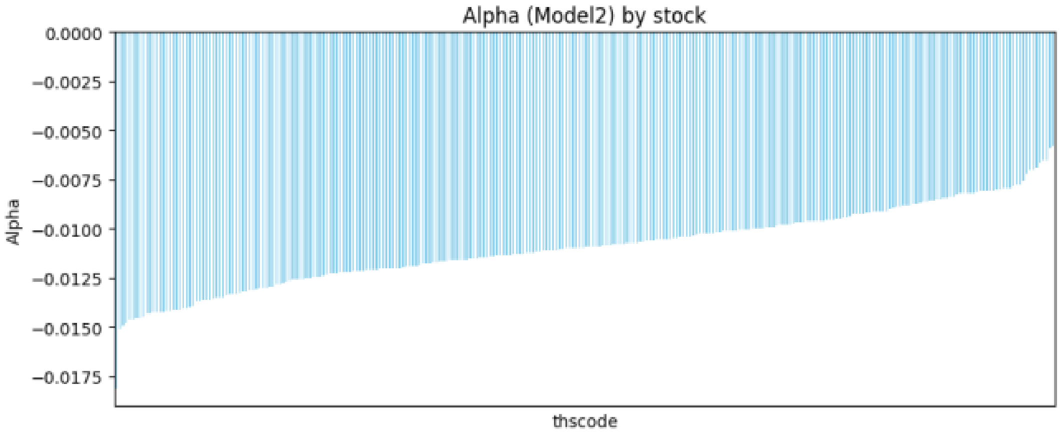

Figure 7. Histogram of intercept values of the expanded five-factor model

3.3.3 Fama-MacBeth Cross-sectional Regression

As shown in Figure 8, the average risk premium of the SENT factor is approximately −1.903 (SE = 0.0607, t ≈ −31.4), indicating that the emotional factor is significantly priced across the cross-section. In contrast, as shown in Figure 9, the risk premiums of each factor in the traditional five-factor model on the left side are all relatively small. The market factor is approximately −0.01598, slightly negative, while the other factors are close to zero and not significant. The traditional factor part of the extended five-factor model on the right side has little change compared to the traditional model, but the SENT factor shows a significant negative premium, with an amplitude much higher than that of other factors. This indicates that the emotional factor is significantly priced across the cross-section, and it does not significantly alter the risk premium structure of the traditional five-factor model.

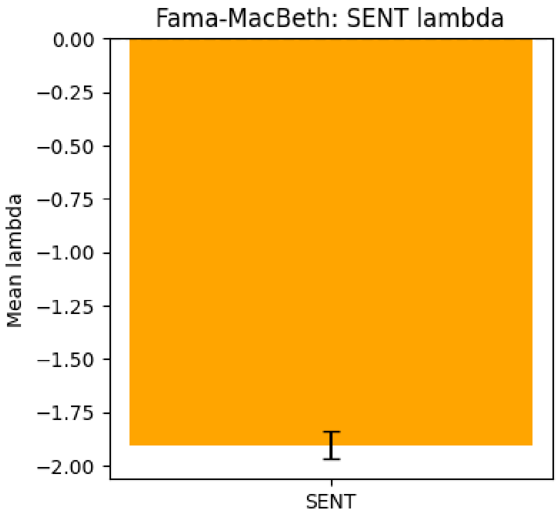

Figure 8. Distribution of the average cross-section risk premium of the sentiment factor.

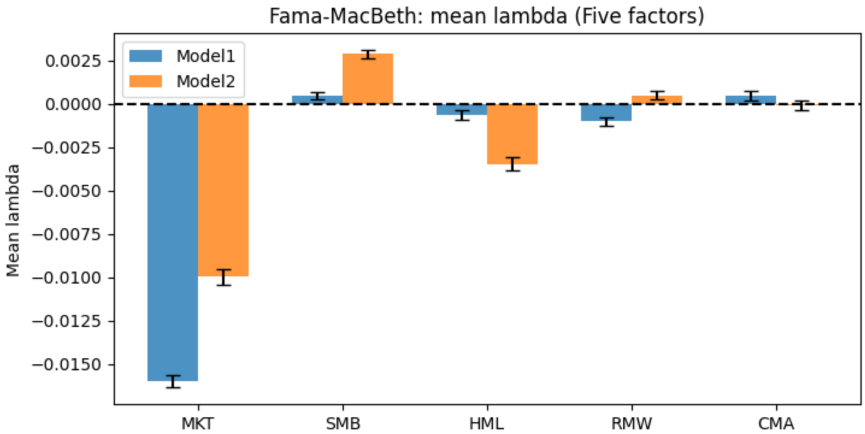

Figure 9. Distribution of the average cross-section risk premia for traditional factors in both the baseline and extended five-factor frameworks.

4. Conclusion

The empirical results show that when the investor sentiment factor (SENT) is incorporated into the traditional five-factor pricing framework, the model's explanation of abnormal returns has significantly improved. The IC and RankIC analyses indicate that the SENT factor has a consistently positive correlation with future returns, suggesting that market sentiment fluctuations can well predict short-term changes in returns. In time series regression, SENT performed significantly on most stocks, indicating its systematic explanatory ability. The Fama-MacBeth regression results show that the average risk premium of this factor is significantly negative, indicating that future returns decline when market sentiment is high. After introducing the sentiment factor, the GRS statistic of the pricing system decreased, indicating that the overall pricing error has significantly converged and the robustness has improved.

This discovery is in line with the viewpoints of behavioral finance. In the A-share market, investors are prone to be influenced by market fluctuations and public opinions, which leads to their emotional trading, amplifying the deviation of prices and making emotions become a systemic risk source in the market. Emotions have a more significant impact on small-cap stocks, high-volatile stocks, and those with difficult arbitrage, which is in line with the conclusion of Baker and Wurgler.

The research results of this paper have significant practical significance. The introduction of investor sentiment factors expands the applicability of traditional models and also verifies the pricing role of this factor in emerging markets, providing empirical support for the application of behavioral finance in asset pricing. Institutional investors can utilize sentiment factors to identify opportunities for excess returns driven by emotions, individual investors can follow this framework to avoid irrational decisions, and regulatory authorities can incorporate it into market monitoring to prevent systemic risks. Future research can further study the time-varying impact of sentiment factors, combine alternative data, and utilize artificial intelligence and machine learning methods to capture the nonlinear relationships between factors, in order to enhance the predictive ability of the pricing system.

References

[1] Carhart, M. M. (1997). On persistence in mutual fund performance. The Journal of Finance, 52(1), 57–82.

[2] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3–56.

[3] Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics, 116(1), 1–22.

[4] Gao, Y., Zhao, C., & Wang, Y. (2024). Investor sentiment and stock returns: New evidence from Chinese carbon-neutral stock markets based on multi-source data. International Review of Economics and Finance, 92, 438–450.

[5] Gniazdowski, Z. (2021). Principal component analysis versus factor analysis.

[6] Habibah, U., Bhayo, M. U. R., & Iqbal, M. S. (2021). Investor sentiments and Fama–French five-factor premia. SAGE Open, 11(2).

[7] Kostin, K. B., Runge, P., & Charifzadeh, M. (2022). An analysis and comparison of multi-factor asset pricing model performance during pandemic situations in developed and emerging markets. Mathematics, 10(1), 142.

[8] Li, Y., & Zhang, Y. (2021). Investor sentiment, idiosyncratic risk, and stock price premium: Evidence from Chinese cross-listed companies. SAGE Open, 11(2).

[9] Markowitz, H. M. (1952). Portfolio selection. The Journal of Finance, 7(1), 77–91.

[10] Titman, S., Wei, C., & Zhao, B. (2022). Corporate actions and the manipulation of retail investors in China: An analysis of stock splits. Journal of Financial Economics, 145(3), 762–787.

[11] Wurgler, J., & Baker, M. (2006). Investor sentiment and the cross-section of stock returns. The Journal of Finance, 61(4), 1645–1680.

[12] Yao, J. (2023). Empirical analysis on construction of mimicking factors and its effect on model adequacy in the Fama-French three-factor model. BCP Business & Management, 38, 164–174.

[13] Zhou, Z. Q., Wu, J. J., Huang, P., & Xiong, X. (2025). Cross-sectional anomalies and conditional asset pricing models based on investor sentiment: Evidence from the Chinese stock market. Financial Innovation, 11, 103.