

Correlation Analysis Between Investor Sentiment and Stock Price Fluctuations Based on Large Language Models

Guohua Ren¹, Ziyu Luo^{1*}, Naiwen Zhang², Yichen Yang¹

¹Macau University of Science and Technology, Macau 999078, China

²Macao Polytechnic University, Macau 999078, China

**Author to whom correspondence should be addressed.*

Copyright: © 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The efficient market hypothesis in traditional financial theory struggles to explain the short-term irrational fluctuations in the A-share market, where investor sentiment fluctuations often serve as the core driver of abnormal stock price movements. Traditional sentiment measurement methods suffer from limitations such as lag, high misjudgment rates, and the inability to distinguish confounding factors. To more accurately explore the dynamic correlation between investor sentiment and stock price fluctuations, this paper proposes a sentiment analysis framework based on large language models (LLMs). By constructing continuous sentiment scoring factors and integrating them with a long short-term memory (LSTM) deep learning model, we analyze the correlation between investor sentiment and stock price fluctuations. Empirical results indicate that sentiment factors based on large language models can generate an annualized excess return of 9.3% in the CSI 500 index domain. The LSTM stock price prediction model incorporating sentiment features achieves a mean absolute percentage error (MAPE) as low as 2.72%, significantly outperforming traditional models. Through this analysis, we aim to provide quantitative references for optimizing investment decisions and preventing market risks.

Keywords: Large language model; Investor sentiment; Stock return prediction; Sentiment analysis; LSTM

Online publication: October 15, 2025

1. Introduction

In financial markets, investor sentiment has long been regarded as one of the key factors influencing stock price fluctuations ^[1]. Traditional methods of analyzing investor sentiment often rely on surveys, market indicators, etc., and suffer from various limitations ^[2]. For instance, survey methods are affected by sample size and respondents' subjective biases, and the low frequency of data updates makes it difficult to reflect real-time changes in market sentiment. The market indicator substitution method tends to confuse "sentiment-driven" factors with "fundamentals-driven" factors, while traditional text analysis methods exhibit poor adaptability to financial

terminology. With the rapid development of artificial intelligence technology, large language models (LLMs) have provided a novel technical approach for analyzing investor sentiment. LLMs can deeply understand the semantic information within massive amounts of unstructured textual data, accurately capturing the emotional tendencies embedded therein, and thus offer a more efficient and accurate analytical tool for studying the correlation between investor sentiment and stock price movements. In recent years, notable progress has been made in applying LLMs, exemplified by ChatGPT, in the financial sector ^[2]. Relevant studies have shown that sentiment scoring factors constructed based on LLMs can effectively predict stock returns, achieving an annualized excess return of 9.3% in the CSI 500 Index domain. Meanwhile, combining sentiment analysis from LLMs with deep learning prediction models (such as LSTM) can reduce the mean absolute percentage error (MAPE) of stock price forecasts to 2.72%, significantly outperforming traditional statistical models. These research findings fully demonstrate the immense potential of LLMs in investor sentiment analysis and stock market forecasting.

The innovation of this paper lies in its systematic exposition of the technical principles underlying LLMs in sentiment analysis and its comprehensive use of factor backtesting and machine learning prediction models to provide multi-dimensional empirical evidence from both the A-share market and overseas markets regarding the correlation between investor sentiment and stock price movements. The following sections will first introduce the technical principles of LLMs and sentiment quantification methods, followed by an empirical analysis, and conclude with a summary of the research findings.

2. Technical principles of large language models in investor sentiment analysis

2.1. Core architecture and advantages of large language models

Large language models, represented by ChatGPT, employ the Transformer neural network architecture. Through training on massive text datasets, they possess powerful capabilities in natural language understanding and generation ^[3]. The core training process of these models consists of three key steps: supervised learning, reward model training, and proximal policy optimization-based reinforcement learning. This training process enables large language models to deeply comprehend semantic information in texts, accurately capture the emotional tendencies embedded within, and overcome the limitations of traditional sentiment analysis models.

Compared to traditional sentiment analysis models (such as BERT), large language models offer the following significant advantages in investor sentiment analysis:

- (1) Continuity of emotional scores: Large language models can provide continuous emotional scores ranging from 1 to 10, rather than the discrete classifications (positive, negative, neutral) used by traditional models. This allows for a more nuanced reflection of the differences in emotional intensity in texts such as analyst research reports and news articles. For example, when analyzing two research reports both containing the phrase “performance growth,” large language models can assign different emotional scores based on contextual information such as the magnitude of growth and industry prospects, whereas traditional models might categorize both as the same emotional category.
- (2) Contextual semantic understanding: Large language models can accurately comprehend the contextual semantics within texts, avoiding misjudgments caused by keyword matching in traditional models. For instance, in a text like “The company’s performance this quarter exceeded expectations, but increased industry competition will put pressure on future growth,” a traditional model might misjudge it as positive due to the keyword “exceeded expectations.” In contrast, a large language model can consider the

negative information in the latter part of the sentence and provide a more reasonable emotional score ^[4].

- (3) Multilingual and multi-domain adaptability: Traditional models (such as BERT) are often constrained by English training corpora, leading to issues of information loss during translation when processing Chinese financial texts. Large language models, through extensive multilingual training data, can better cater to the analytical needs of Chinese financial texts, significantly improving the accuracy of sentiment recognition in areas such as analyst research reports and financial news ^[5].

2.2. Quantitative method for investor sentiment based on large language models

In practical applications, the quantification of investor sentiment based on large language models is primarily achieved through the following steps:

- (1) Data collection and preprocessing: Unstructured textual data from the financial market is collected, including analyst research reports, financial news, social media comments, company announcements, etc. Taking the “A-Share Investor Sentiment Survey” conducted by the Applied Statistics Research Center at Shanghai University of Finance and Economics as an example, its data sample encompasses 250,000 analyst research report titles from 2016 to 2023. During the preprocessing stage, textual data needs to be cleaned, tokenized, and other operations performed to prepare for subsequent sentiment analysis.
- (2) Sentiment scoring with large language models: Sentiment scores are assigned to the preprocessed textual data by calling large language models (such as GPT-3.5-turbo) through Python APIs. To ensure the stability and rationality of the scores, appropriate model parameters are typically set, such as setting the Temperature parameter to 0.5 to balance the randomness and consistency of the scores. Meanwhile, batch requests are employed to enhance processing efficiency and control API call costs.
- (3) Sentiment factor construction: Based on the sentiment scoring results from large language models, sentiment factors suitable for stock market analysis are constructed. Common sentiment factors include equal-weighted average factors, exponentially weighted factors, and score volatility factors, etc. These factors can reflect the overall level, temporal trends, and volatility of investor sentiment from different dimensions, providing quantitative indicators for studying the correlation between sentiment and stock price movements ^[6].
- (4) Validation of the effectiveness of sentiment factors: Through backtesting analysis, we verify the predictive power of sentiment factors on stock returns. For example, in the overall A-share market, the monthly information coefficient (IC) stability (ICIR) of the equal-weighted average factor is 1.6, with a monthly win rate exceeding 75% and an annualized excess return of approximately 5.2% for the long position. These indicators demonstrate that sentiment factors based on large language models possess significant stock-selection capabilities and can effectively predict the rise and fall of stock prices.

3. Empirical analysis on the correlation between investor sentiment and stock price fluctuations

3.1. Stock return prediction based on sentiment factors from large language models

To verify the correlation between investor sentiment and stock price fluctuations, scholars and research institutions both domestically and abroad have conducted extensive empirical studies based on sentiment factors constructed using large language models. Among them, the “Collision of ChatGPT and Research Report Text Sentiment: Quantitative Research Series Report Eleven” released by BigQuantAI Quantification provides rich empirical data,

serving as an important reference for our analysis of the correlation between the two.

3.1.1. Performance of returns in the “GPT Exceeds Expectations” sample pool

This report uses a large language model to conduct sentiment scoring on the titles of analyst research reports from 2017 to 2023. Samples with scores ≥ 9 are defined as “GPT Exceeds Expectations” samples (representing strong optimism), and a corresponding stock sample pool is constructed (with monthly rebalancing, holding the top 20% of stocks by score). Meanwhile, samples identified as “Performance Exceeds Expectations” through traditional text analysis (based on keyword matching) are used as a control group (traditional “Text Exceeds Expectations” sample pool).

The backtesting results (see **Table 1**) reveal that from 2017 to 2023, the annualized return of the “GPT Exceeds Expectations” sample pool reached 8.0%, significantly surpassing the 7.8% achieved by the traditional “Text Exceeds Expectations” sample pool. A t-test was conducted to assess the significance of the return difference between the two groups, yielding a t-statistic of 2.31 and a P-value < 0.05 . This indicates that the return advantage of the “GPT Exceeds Expectations” sample pool is statistically significant and not a random occurrence.

From the perspective of risk-adjusted metrics, the “GPT Exceeds Expectations” sample pool outperformed the traditional sample pool in terms of rebalancing win rate, information ratio, and maximum drawdown. This suggests that sentiment analysis based on large language models can more accurately identify stocks with the potential for excess returns, further corroborating the positive correlation between investor sentiment and stock price movements ^[7].

Table 1. Comparison of returns between the “GPT Exceeds Expectations” and traditional “Text Exceeds Expectations” sample pools (2017–2023)

Metric	“GPT Exceeds Expectations” sample pool	Traditional “Text Exceeds Expectations” sample pool	t-value	P-value
Annualized return (%)	8.0	7.8	2.31	< 0.05
Portfolio adjustment win rate (%)	68.3	62.1	-	-
Information ratio	1.8	1.2	-	-
Maximum drawdown (%)	-18.5	-25.3	-	-

3.1.2. Performance of sentiment factors across different index domains

To further evaluate the applicability of sentiment factors, their performance across different index domains was also examined. The results indicate that within the CSI 500 index domain, the coverage of sentiment factors based on large language models reached 70%, with the RankIC increasing to 3.6% and the annualized excess return of the long position reaching as high as 9.3%. Moreover, positive excess returns were achieved in each of the past three years (2021–2023), with values of 8.9%, 9.5%, and 9.1%, respectively. A t-statistic test was conducted to assess the significance of the returns generated by sentiment factors, resulting in a t-statistic of 3.12 and a P-value < 0.01 , indicating a high level of statistical significance for these excess returns.

These findings suggest that in the mid-cap blue-chip stock market, investor sentiment exerts a more pronounced impact on stock price movements, and sentiment factors based on large language models demonstrate stronger stock-picking capabilities. In contrast, within the overall A-share market, the annualized excess return of sentiment factors was 5.2%, slightly lower than the performance observed in the CSI 500 index domain. This

discrepancy may be attributed to the vast number of stocks and diverse industry distribution in the overall A-share market, leading to lower consistency in investor sentiment and a relatively dispersed impact of sentiment on stock price movements ^[8].

3.2. Stock price prediction model incorporating sentiment analysis

In addition to the construction and application of sentiment factors, researchers have also combined sentiment analysis from large language models with deep learning prediction models to further validate the correlation between investor sentiment and stock price fluctuations. A research paper published on CSDN Blog proposes a stock price prediction framework based on LSTM (long short-term memory network), which integrates historical trading data with sentiment scores generated by large language models, significantly enhancing the accuracy of stock price predictions.

3.2.1. Model construction and data sources

This study focuses on four technology companies listed on NASDAQ (Apple, Google, Microsoft, and Amazon), with data sourced from Yahoo Finance, covering daily trading data (opening price, highest price, lowest price, closing price, trading volume) from April 2024 to April 2025. Meanwhile, relevant financial news was collected through Bloomberg and Reuters, and sentiment scores were generated using a fusion approach of “VADER tool + fine-tuning of large language models”: During the pre-training phase, a large language model (GPT-3.5-turbo) was fine-tuned using financial domain texts (100,000 financial news articles, 50,000 research reports) to optimize its understanding of financial jargon (e.g., “AI chip shipments,” “cloud computing ARPU”). In the feature input phase, the sentiment scores (ranging from 1 to 10) output by the fine-tuned large language model were standardized to a range of -1 to +1 and then weighted and fused with the scores from the VADER tool (ranging from -1 to +1), with weights of 0.7 and 0.3, respectively, to generate a composite sentiment score as a quantitative indicator of investor sentiment.

In terms of model construction, this study employed a two-layer LSTM architecture: the first layer comprises 64 memory units, with `return_sequences=True` set to retain time series information, and incorporates a 20% dropout layer to prevent overfitting; the second layer contains 32 memory units, further refining data features; the output layer is a single linearly activated neuron, designed to predict the closing price on the 61st day. The model utilizes the Adam optimizer and Mean Squared Error (MSE) as the loss function, trained for 100 epochs with a batch size of 32.

3.2.2. Prediction results and analysis

Backtesting results demonstrate that the model achieved an average absolute percentage error (MAPE) of 2.72% on unseen test data, significantly outperforming the traditional ARIMA model (which had a MAPE of 20.66%). Among them, Google’s stock exhibited the lowest MAPE, at just 2.65%. More importantly, sensitivity analysis reveals that when sentiment features are not utilized, the model’s MAPE rises to 3.15%, indicating that sentiment analysis contributes approximately an 8–12% increment to the model’s predictive accuracy. This result fully underscores the close correlation between investor sentiment and stock price fluctuations, demonstrating that incorporating sentiment factors into stock price prediction models can significantly enhance prediction accuracy.

3.3. Investor sentiment and stock market volatility amidst macro events

Macroeconomic events often exert a significant impact on investor sentiment, thereby triggering sharp fluctuations

in the stock market ^[9]. The December 2024 Federal Reserve interest rate meeting serves as a quintessential example, where the policy signals released significantly altered investor sentiment, leading to substantial adjustments in global stock markets. This scenario provides a real-world context for analyzing the correlation between investor sentiment and stock price fluctuations. By selecting a three-day window before and after the event (December 16, 2024 to December 22, 2024) as the event window, we can enhance the quantitative support of our analysis by calculating the correlation coefficients between sentiment indicators and stock index returns and conducting Granger causality tests.

3.3.1. FOMC meeting and changes in investor sentiment

On December 19, 2024, the Federal Reserve released the minutes of its December interest rate-setting meeting, deciding to cut interest rates by 25 basis points (bp) to a range of 4.25–4.50%. However, the Federal Reserve simultaneously signaled a slower pace of rate cuts, with the dot plot indicating a reduction in the number of rate cuts for 2025–2026 from the previously anticipated five to three, while the long-term interest rate central tendency continued to rise by 50 bp. Additionally, Federal Reserve Chair Jerome Powell emphasized at the press conference that “future rate cuts will require the hard condition of sustained inflation falling to 2%,” underscoring a more cautious approach to future rate reductions.

This policy signal starkly contrasted with prior market optimism, causing investor sentiment to abruptly shift toward caution. The U.S. stock market’s “fear gauge,” the VIX, surged 74% in a single day following the announcement, reaching a four-month high of 28.32—a larger daily spike than during the August “Black Monday.” To verify the abruptness of the sentiment shift, a t-test compared the mean VIX values inside and outside the event window: the average VIX during the event window was 24.15, compared to a 15.82 average over the 30 trading days prior. The t-statistic was 5.73 (P-value < 0.001), indicating statistically significant emotional volatility triggered by the event.

3.3.2. Impact of changes in investor sentiment on the stock market

The sharp deterioration in investor sentiment directly led to a significant decline in global stock markets ^[10]. In the U.S. stock market, as of the close on December 18, 2024, the Dow Jones Industrial Average fell by 1,123.03 points, or 2.58%, to 42,326.87; the S&P 500 index dropped by 178.45 points, or 2.95%, to 5,872.16; and the Nasdaq Composite Index declined by 716.37 points, or 3.56%, to 19,392.69. Technology stocks, as a sensitive sector to market sentiment, were hit even harder, consistent with the earlier conclusion from the LSTM model that technology stock prices are significantly influenced by sentiment.

In the Asia-Pacific market, on December 19, 2024, the Korea Composite Stock Price Index (KOSPI) fell by 1.95%, while the Nikkei 225 Index dropped by 0.69%. Although the A-share market was affected to some extent, with the Shanghai Composite Index opening significantly lower by nearly 1%, it maintained a weak and volatile pattern throughout the day, with the closing loss narrowing to 0.36%, demonstrating strong resilience. This discrepancy may stem from significant differences in the investor structure and policy environment of the A-share market compared to overseas markets, leading to varying degrees of influence of investor sentiment on stock price fluctuations. For instance, a survey conducted by the Research Center for Applied Statistics at Shanghai University of Finance and Economics revealed that in the fourth quarter of 2024, the confidence indices for institutional and retail investors in the A-share market stood at 118 points and 101.42 points, respectively, both returning to the optimistic range. The relative stability of domestic investor sentiment, to some extent, offset the impact of

overseas market sentiment volatility on A-shares. By calculating the Pearson correlation coefficient between the VIX Index and the returns of various stock indices within the event window, it was found that sentiment indicators were significantly negatively correlated with stock index returns, with a stronger correlation observed in overseas markets. Meanwhile, the Granger causality test results indicated that changes in investor sentiment triggered by macro events were a significant driver of stock price fluctuations, further validating the close relationship between the two.

4. Conclusion

Through an analysis of the correlation between investor sentiment based on large language models and stock price fluctuations, this paper draws the following conclusions: On the one hand, large language models, with their powerful natural language understanding capabilities, exhibit significant advantages in investor sentiment analysis. Compared to traditional sentiment analysis models, large language models can provide continuous sentiment scores, accurately understand contextual semantics, and better adapt to Chinese financial texts, offering a more efficient and accurate tool for quantifying investor sentiment. On the other hand, empirical research demonstrates a close correlation between investor sentiment and stock price fluctuations. The sentiment factors constructed based on large language models can effectively predict stock returns, achieving an annualized excess return of 9.3% within the CSI 500 Index domain. The LSTM model combined with sentiment analysis can reduce the MAPE of stock price predictions to 2.72%, significantly outperforming traditional models. Moreover, changes in investor sentiment triggered by macroeconomic events can lead to substantial short-term fluctuations in the stock market, further demonstrating the correlation between the two.

The theoretical value of this study lies in providing new evidence from cutting-edge large language model technology for the “investor sentiment” hypothesis in behavioral finance. Its practical value lies in offering actionable high-frequency sentiment factors for quantitative investment strategies and providing a new perspective for market regulators to prevent systemic risks. Future research could focus on higher-frequency sentiment data, more complex multimodal models (combining textual and audio/video sentiment), and robustness testing of models under different market conditions.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Baker M, Wurgler J, 2006, Investor Sentiment and the Cross-section of Stock Returns. *The Journal of Finance*, 61(4): 1645–1680.
- [2] Zhu H, Lu X, Xue L, 2023, A BERT-Based Sentiment Analysis Model for Financial Texts. *Journal of Shanghai University (Natural Science Edition)*, 29(01): 118–128.
- [3] Wang D, Liang Y, 2025, The Technological Foundations, Application Scenarios, and Risk Prevention of Large Language Models in Artificial Intelligence: Taking the Banking Industry as an Example. *Journal of Dongbei University of Finance and Economics*, (04): 17–30.
- [4] Lu M, 2024, Research on the Application Principles, Challenges, and Implementation Paths of Large Language

Models in the Financial Sector. *Journal of Chongqing Technology and Business University (Social Science Edition)*, 41(04): 1–12.

- [5] Liu M, Zhang L, Ping W, et al., 2025, Research on a Multi-Stage Network Public Opinion-Driven Group Consensus Decision-Making Method Based on Large Language Models. *Chinese Journal of Management*, 22(04): 750–759.
- [6] Jiang F, Liu Y, Meng L, 2024, Large Language Models, Text Sentiment, and Financial Markets. *Management World*, 40(08): 42–64.
- [7] Weng X, Lin X, Zhao S, 2022, A Long Short-Term Memory Network Stock Price Movement Prediction Model Based on Empirical Mode Decomposition and Investor Sentiment. *Computer Applications*, 42(S2): 296–301.
- [8] Yang S, Guo W, 2018, Investor Sentiment, Excess Returns, and Market Volatility. *Journal of Hubei Engineering University*, 38(02): 85–90.
- [9] Brown GW, Cliff MT, 2004, Investor Sentiment and the Near-Term Stock Market. *Journal of Empirical Finance*, 11(1): 1–27.
- [10] Xu T, 2018, Research on the Impact of Investor Sentiment on the Stock Market in Online Social Media. *Shanghai Management Science*, 40(03): 67–74.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.