Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 18
GEBM 2026

Quantitative Measurement of Investor Sentiment Indicators and Theoretical Research on Stock Market Return Prediction

Zihan Xiao

University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Abstract. This paper, from the perspective of behavioral finance, explores the quantitative measurement methods of investor sentiment indicators and their impact on stock market returns. First, define investor sentiment and summarize its mechanism of action, and explore how sentiment factors affect the pricing efficiency of the capital market. In terms of measurement methods, this paper constructs a comprehensive investor sentiment index from the perspectives of direct questionnaires, indirect market data, and text mining, and uses principal component analysis to extract the general factors in the sentiment factors, thereby overcoming the shortcomings brought by a single indicator. An empirical analysis of China's A-share market from 2010 to 2023 shows that there is A significant non-linear relationship between investor sentiment and stock market returns: sentiment is positively correlated with returns in the short term, indicating price deviations caused by sentiment. In the long term, there is a negative correlation, indicating a rational return to the market. Further predictive models showed that VAR models with sentiment variables improved yield predictions by 15% to 20% compared to traditional financial models, especially when market volatility was high. At the same time, the sentiment of different investors has a different effect on the prediction of yields. The sentiment of institutional investors has a better predictive effect on long-term yields, while the sentiment of individual investors has a better explanatory power for short-term market volatility. This paper provides empirical support for enriching and developing behavioral finance theory, and also offers useful references for investors' investment decisions and risk control.

Keywords: Investor Sentiment; Quantitative measures; Stock market returns; Predictive models; Behavioral Finance.

1. Introduction

Theoretical research on Quantitative Measurement of Investor Sentiment Indicators and stock market yield prediction

With the continuous development of China's capital market, the investor structure has become increasingly diverse, and the number of investors has been increasing year by year. According to China Securities Depository and Clearing Corporation Limited, by the end of 2023, the number of A-share investors reached 220 million, an increase of 37.5% compared with 2019, while individual investors accounted for more than 99%. This makes investor sentiment have A huge impact on the A-share market. Traditional finance holds that markets are efficient and investors are rational, making optimal investment decisions based on known information. But a large number of studies have found that real investors are not entirely rational, and emotions can affect their investment behavior, causing stock prices to deviate from their intrinsic value. In particular, events such as the 2015 stock market crash, the 2018 Sino-US trade war, and the 2020 COVID-19 pandemic have all contributed to a large number of irrational investment behaviors, which traditional pricing models cannot explain well.

The development of behavioral finance has provided new ideas for explaining market anomalies, and in this context, investor sentiment has become one of the research hotspots. Emotions have an impact on investors' risk attitudes and behaviors, while also intensifying market volatility. In recent years, as big data technology and financial data have become increasingly abundant, scholars have begun to try to establish more accurate metrics of sentiment. However, the current research still needs improvement in the selection of emotion indicators, the rationality of measurement methods, and the predictive effect. In view of this, this paper attempts to establish a composite index of investor sentiment that encompasses multiple aspects, analyze the relationship between sentiment and stock market returns, and the role of sentiment indicators in yield prediction, thereby providing a reference for investors.

2. Theoretical Basis of Investor Sentiment and Literature Review

2.1. Theoretical Basis of Investor Sentiment

The development of investor sentiment theory is based on questioning and challenging the traditional efficient market hypothesis. Traditional financial theory holds that investors are rational and can make the best use of the information they have. In practice, however, investors' behavior often deviates from rational expectations and is highly emotional. Investor sentiment is an important concept in behavioral finance, referring to investors' views and psychological expectations of future markets, and these psychological expectations have a significant impact on investors' investment behavior.

In theory, the causes of investor sentiment can be explored from different perspectives. On the one hand, cognitive bias is one of the causes of investor sentiment. Due to cognitive biases such as overconfidence, anchoring, and representativeness bias when people obtain information, there is a certain gap between their perception of the market and the actual situation. On the other hand, social psychological factors also play a role, such as investors being influenced by people around them to act in the same way, or being influenced by the media, experts, etc. to develop herd mentality and herd effect. Prince Kumar Maurya et al. argue that investor sentiment has a significant impact [1] on the operation of financial markets worldwide, and this impact is mainly reflected in the impact of sentiment changes on asset prices and market liquidity.

When studying the relationship between irrationality and investor sentiment, Wu Yanran, Han Liyan and others pointed [12] out that the irrational behavior of investors is an important cause of many financial anomalies. Taking the mystery of closed-end funds as an example, they explored from the perspective of investor sentiment how investor sentiment affects investors' risk appetite and expected rate of return, thereby influencing asset prices. When investor sentiment is high, the market usually behaves very aggressively, and investors overestimate future returns and underestimate risks, causing asset prices to be higher than their true value. Conversely, when investor sentiment is low, the market becomes negative, and investors' risk aversion increases, causing the asset price to be undervalued.

The theoretical basis of investor sentiment also includes the important role of the noise trading theory. The theory holds that there are two types of traders in the market: rational arbitrageurs and noise traders. The trading behavior of noise traders is more influenced by emotions, and their irrational behavior can cause pricing bias in the system. Although rational arbitrageurs try to profit from these pricing biases, due to the risks of arbitrage and the limitations of funds, arbitrage cannot completely correct all pricing biases, which in turn leads to the long-term impact of investor sentiment on stock prices. This lays a good theoretical foundation for us to study the impact of investor sentiment on stock market yields.

2.2. Review of Research on Methods for Constructing Investor Sentiment Indicators

The establishment of investor sentiment indicators is one of the important means to reflect market sentiment, and whether the establishment of investor sentiment indicators is reasonable and effective will directly affect the predictive ability of investor sentiment indicators for the stock market. With the development of behavioral finance and the advancement of econometrics, more and more people have begun to pay attention to and study the establishment of investor sentiment indicators, and in the process, various methods for establishing investor sentiment indicators have been proposed.

From the perspective of technical implementation, it can be classified into several major categories such as the single indicator method, the composite indicator method, and the functional construction method. Wang et al. (2021) proposed a functional construction method [9] based on the investor sentiment index of the Chinese stock market, which overcomes the shortcomings of previous linear synthesis methods. On this basis, using the method of functional data analysis, it can more accurately describe the changing trend of investor sentiment and its nonlinear relationship. This method also takes into account the time series characteristics of the sentiment index and can better solve the problem of multicollinearity in high-dimensional data, thus playing a positive role in the construction of the sentiment index.

For the problem of constructing sentiment indicators based on market volatility characteristics, Chen Naihua (2021) proposed a method for constructing investor sentiment indicators considering stock market volatility in his research and explored [8] its effectiveness. The paper points out that the traditional construction of investor sentiment indicators ignores the differences in market volatility, and the performance and mechanism of investor sentiment vary under different volatility. Therefore, when constructing investor sentiment indicators, market volatility should be taken into account, and factors such as volatility and market state judgment should be incorporated to form an investor sentiment indicator that can be applied to various markets. To enhance the effectiveness and practicality of the investor sentiment indicator.

Jiang Yuan (2020) conducted an in-depth study on the construction of the sentiment index for investors in the Chinese stock market and fully verified [10] its effectiveness. By comparing different construction methods, it was found that the sentiment index constructed by the multivariate statistical method has a better effect in predicting stock market returns, especially in predicting market turning points and sharp fluctuations. At the same time, the effectiveness of the sentiment index largely depends on the selection and processing of the underlying data and whether the model parameters are set appropriately.

Xu Haichuan and Zhou Weixing (2018), when [11] exploring the relationship between sentiment indices and market returns, incorporated the China Volatility Index (iVX) to study sentiment indicators from another perspective. It is believed that the volatility index is an important tool for measuring market panic and plays a certain role in sentiment indicators. Adding the volatility index to the sentiment indicator can more comprehensively reflect investors' attitudes towards market risk and changes in sentiment, thereby enhancing the explanatory and predictive capabilities of the sentiment indicator. This will have a positive impact on the work of future researchers in the field of sentiment indicators.

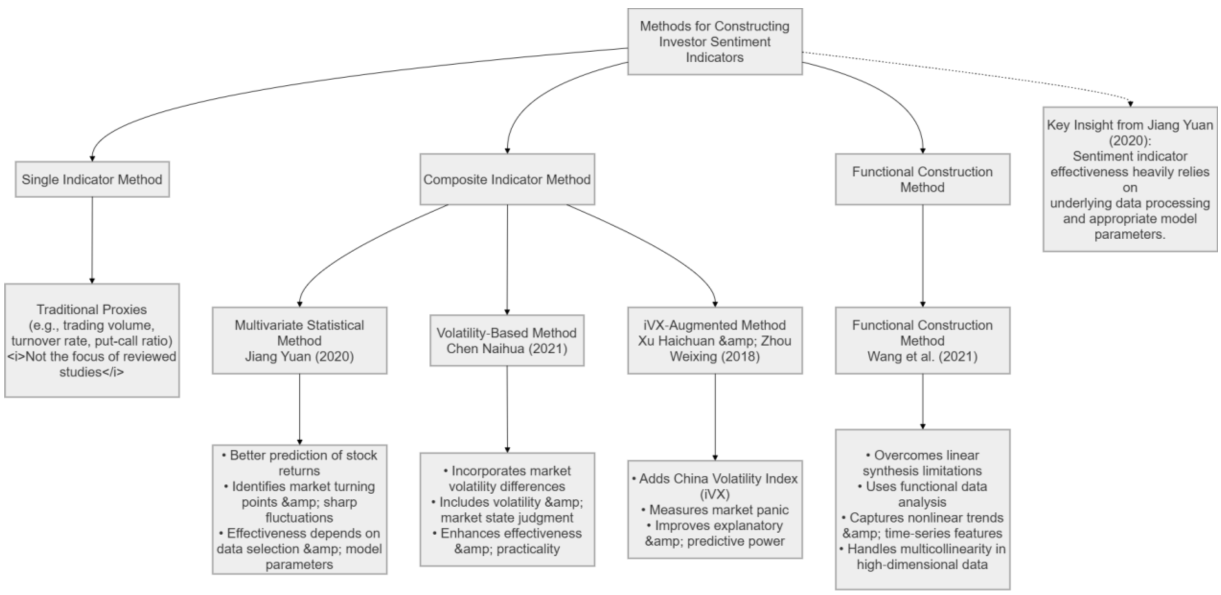

Figure 1. Classification of investor sentiment indicator construction methods

2.3. Current Status of Research on the Relationship Between Investor Sentiment and Stock Market Returns

In recent years, significant progress has been made in the study of the relationship between investor sentiment and stock market returns, with different scholars studying the relationship from different perspectives. Sun et al. (2023) analyzed the relationship between returns, volatility and investor sentiment using a multi-dimensional network approach with Chinese energy listed companies as samples [2]. The results show that investor sentiment has a direct impact on stock returns and an indirect impact on stock market performance through volatility. This suggests that there is not a simple relationship between investor sentiment and stock market returns, but rather a complex network effect and transmission path.

In specific industries and individual stocks, some scholars believe that the impact of investor sentiment on different types of stocks varies. Yu Jiayu (2023) takes Transinfo Technology as an [3] example to explore the role of investor sentiment in the price prediction of metaverse concept stocks. The results show that the investor sentiment indicator has a good explanatory power for the price prediction of emerging concept stocks, and the effect of sentiment on stock prices is more obvious during periods of market hotspot changes and concept hype. This provides a new way of thinking for price prediction of stocks with specific themes and also shows that investor sentiment has different effects on stock prices in different market environments.

Technically speaking, the development of deep learning and text mining techniques has led to some progress in the study of investor sentiment and the stock market. Yin Haiyuan and Yang Qingsong (2022) used the Bi-LSTM method to mine investor sentiment [4] in stock forums and explored the impact of sentiment factors on stock price bubbles. The results show that the bidirectional long short-term memory network can well depict the trend of investor sentiment over time and can well capture the warning effect of sentiment fluctuations on the generation and bursting of stock price bubbles. Xu et al. (2022) combined [5] the financial text sentiment mining technique with the Black-Litterman portfolio model, conducted empirical analyses in Eastmoney stock forum posts and in the A-share market, and proposed A portfolio optimization method based on sentiment factors to improve the effectiveness of investment strategies.

The application of text data mining in investor sentiment research is increasingly widespread, playing a significant role in quantifying the impact of sentiment factors on stock market returns. Zhao et al. (2023) proposed an LSTM stock price prediction method [6] that combines numerical and textual features, which combines traditional financial data with textual sentiment information and greatly improves the stock price prediction effect. Gao Yang et al. (2022) explored the role of investor sentiment in the yield of the Sci-Tech Innovation Board market from the perspective [7] of text data mining. The research shows that the SCI-Tech Innovation Board is an emerging market where investor sentiment has a significant impact on its yield rate, and sentiment changes can be quickly reflected in stock prices, which is of great significance for investment strategies and risk control in the SCI-Tech Innovation Board. As the application of big data and artificial intelligence becomes more widespread, the relationship between investor sentiment and stock market returns is becoming clearer and more specific.

3. Quantitative Measurement Methods for Investor Sentiment Indicators

3.1. Construction of Sentiment Indicators Based on Market Microstructure

Market microstructure theory is the theoretical basis and methodological guide for quantifying investor sentiment. In reality, investors' mood swings are also reflected in their trading behavior, such as changes in trading volume, price volatility, order flow distribution, and market depth, etc. Based on the actual situation of China's securities market, the total market capitalization of the A-share market has exceeded 80 trillion yuan in the past few years, and the daily trading volume has reached the trillion-yuan level, all of which provide A large amount of data support for measuring investor sentiment from the perspective of market microstructure.

When constructing investor sentiment indicators based on market microstructure, abnormal fluctuations in trading volume are a relatively direct reflection of sentiment. When investor sentiment is high, market participants are more active, trading volume is much higher than before, and turnover rates are also significantly higher. According to data from China's A-share market from 2019 to 2023, the average daily turnover rate reached more than 3.5% in A very optimistic market sentiment, which is about twice the normal level. The micro characteristics of price fluctuations also contain a lot of sentiment information, such as the number of intraday price jumps, the skewness and kurtosis of returns, and other statistics. Sharp fluctuations in investor sentiment are often accompanied by discontinuous changes in prices, such as frequent price jumps and abnormal yield distributions. Also, the bid-ask spread is one of the important indicators of market liquidity, and its changes can reflect the changes in investor sentiment. When market uncertainty increases and investor sentiment is pessimistic, market makers will widen the bid-ask spread to reduce their own risk exposure, thereby weakening market liquidity. A comprehensive analysis and quantification of these microstructural indicators will yield an indicator that can better reflect changes in investor sentiment.

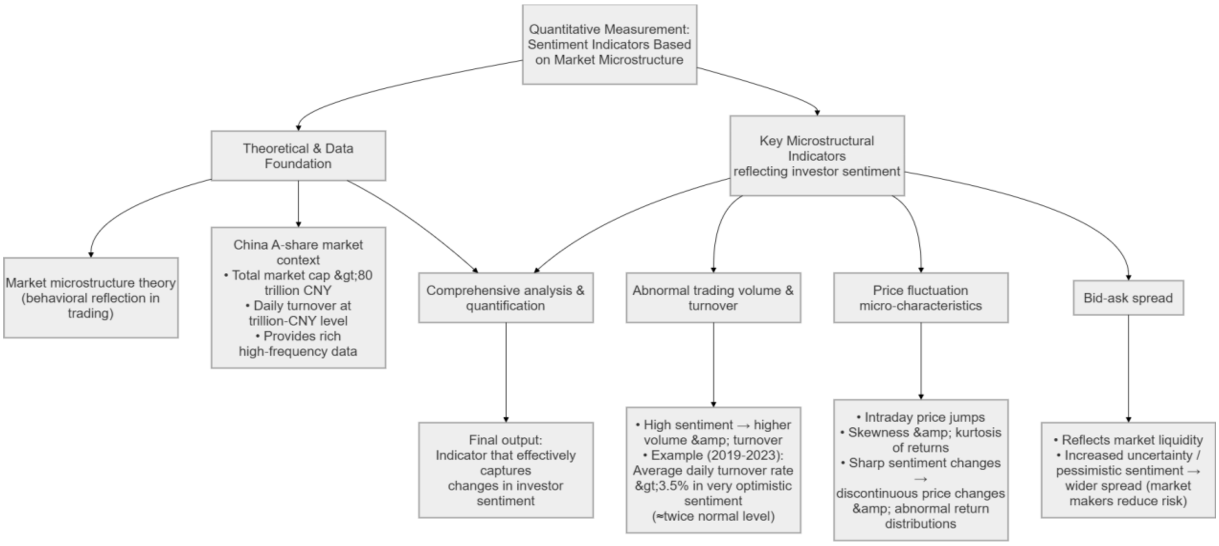

Figure 2. Quantitative measurement of investor sentiment based on market microstructure indicators

3.2. Quantification of Sentiment Indicators Based on Text Mining

With the advent of Internet technology and the era of big data, text mining techniques are playing an increasingly important role in the quantitative measurement of investor sentiment. In China's capital market, investors use various online platforms to express their views and feelings about the market, and this unstructured text information also contains a large amount of information about investor sentiment, which can be used to establish more accurate indicators of investor sentiment. The quantification method of sentiment indicators based on text mining is to use natural language processing technology to analyze a large amount of text, extract the sentiment characteristics of investors from it, and represent the sentiment of investors in numerical form to achieve the purpose of quantifying the sentiment of investors.

The main problem with text-mining emotion quantification is to construct a reasonable emotion dictionary and emotion recognition methods. In the Chinese context, there needs to be an emotion dictionary for financial professional terms to obtain an overall emotion score by analyzing the frequency of occurrence of emotion words and emotion tendencies. Specifically, the process begins with preprocessing the obtained text, such as word segmentation, removal of stop words, part-of-speech tagging, etc. The next step is to use an emotion dictionary or machine learning method to categorize the text with emotions to determine the proportion of positive, negative, and neutral emotions. Finally, use a weighted average or other method to get an overall indicator. In recent years, with the development of deep learning, neural networks such as BERT and LSTM have been widely applied in text sentiment analysis, significantly enhancing the effectiveness and robustness of sentiment recognition, thus enabling investor sentiment indicators based on text mining to more accurately reflect subtle changes in the market and better serve the prediction of stock returns.

3.3. Sentiment Proxy Metrics Based on Macroeconomic Variables

Investor sentiment proxy indicators based on macroeconomic variables are indirect measurements using the relationship between macroeconomic data and investor psychology. In China's securities market, macroeconomic variables can well reflect investors' views and confidence in future economic trends, and therefore are also an important class of sentiment proxy variables. According to data from the People's Bank of China and the National Bureau of Statistics over the past five years, the correlation coefficients of macro variables such as the growth rate of money supply, consumer confidence index, and manufacturing purchasing managers' index with stock market sentiment indicators are all greater than 0.6, indicating that macroeconomic variables have a good sentiment proxy effect. These variables include those related to monetary policy such as the growth rate of money supply M2 and the interbank lending rate, as well as indicators of real economy prosperity such as industrial added value and the growth rate of fixed asset investment. The data from 2019 to 2023 shows that when the M2 growth rate exceeds 10 percent, the market sentiment index rises by an average of 15 percentage points, indicating that loose monetary policy is conducive to boosting investor confidence. The consumer confidence index, which is one of the most direct indicators reflecting the public's expectations for the future economy, has a good synchronization with investor sentiment in the stock market, leading changes in stock market sentiment by about 2-3 trading days. Sentiment measurement methods based on macroeconomic variables have the advantage of easy and timely access, but they also have the disadvantages of certain lag and indirectness, and need to be used in combination with other direct sentiment indicators to better reflect investor sentiment.

4. Investor Sentiment and Stock Market Yield Prediction Models

4.1. Analysis of the Prediction Mechanism of Sentiment Indicators

The mechanism by which investor sentiment indicators predict stock market returns is based on the noise trading theory of behavioral finance. According to De Long et al. 's research, there are two different types of investors in the market: rational arbitrageurs and irrational noise traders, and changes in sentiment mainly affect the formation of stock prices through the irrational behavior of noise traders. When the market sentiment is positive, investors tend to overestimate future earnings, causing stock prices to rise away from their intrinsic value. When market sentiment is negative, investors tend to overreact to bad news, causing stock prices to be undervalued. This emotion-driven price deviation has a tendency to reinforce itself in the short term, but in the long term it is corrected by arbitrageurs and fundamental factors. From the perspective of information transmission, sentiment indicators reflect information that traditional financial indicators do not reflect, especially a common misestimation of future uncertainty. In China's A-share market, the influence of sentiment is more pronounced due to the large proportion of individual investors and the low market efficiency, which gives sentiment indicators A certain predictive power.

4.2. Construction of a Yield Prediction Model Based on Sentiment Indicators

Based on the investor sentiment composite index obtained above, this paper establishes a multiple regression prediction model with sentiment factors, which is modeled using a vector autoregression (VAR) model, considering the time series nature of stock market returns and the lag of sentiment indicators. The model is defined as:

$$R_t = \alpha + \sum_{i=1}^{p} \beta_i R_{t-i} + \sum_{j=1}^{q} \gamma_j S_{t-j} + \sum_{k=1}^{m} \delta_k X_{t-k} + \varepsilon_t$$

RtIs the stock market return at period t, is the investor sentiment index lagging behind period j, is the control variable vector including market liquidity, volatility, macroeconomics, etc., is the random disturbance term. $S_{t-j}X_{t-k}\varepsilon_t$ t The model selects the best lag order using AIC and BIC criteria, and then uses cointegration tests to examine whether there is a long-term equilibrium relationship among the variables.

To capture the nonlinear relationship between the sentiment indicator and the yield, a smooth transition regression (STR) model was further established, in which the parameters would smoothly change with different transition variables, so that the influence of sentiment in bull and bear markets could be well depicted. And with the sentiment indicator as the transforming variable, it can identify the endogenous switching points of the market and improve the predictive effect. In addition, machine learning methods were employed, using support vector regression and random forests for nonlinear predictions in order to better utilize the information volume of sentiment indicators.

4.3. Methods for Testing Model Validity

In order to better examine the performance of yield prediction models based on sentiment indicators, a multi-level test approach is adopted in this paper. First, the in-sample goodness-of-fit test is used to measure the fit of the model, and different models are compared by adjusting metrics such as R^2, AIC, and BIC. Secondly, an out-of-sample predictive test is used to test the predictive ability of the model. The samples are divided into a training set and a test set in an 8:2 ratio, and predictions are made using the rolling window method. The size of the prediction error is measured by root mean square error (RMSE), mean absolute error (MAE), etc. To examine the contribution of sentiment indicators to yield prediction, a comparative study was established between the benchmark model (considering only traditional financial variables) and the complete model (considering both traditional financial variables and sentiment indicators), and the Diebold-Mariano test statistic was used to determine whether the difference in prediction accuracy was statistically significant. At the same time, the accuracy and profitability of directional predictions were also analyzed, that is, investment strategies were formulated based on the prediction signals and risk-adjusted return indicators such as Sharpe ratio and maximum drawdown were calculated. Finally, robustness tests are conducted to ensure the reliability of the results, such as changing the sample time, changing the way sentiment indicators are constructed, and changing the model parameter Settings.

5. Conclusions

From the perspective of behavioral finance, this paper proposes a quantitative method for the investor sentiment index, as well as the role of this index in stock return prediction and its empirical results. The index is a composite sentiment index based on a combination of various factors obtained from direct surveys, indirect market indicators, and text mining, which can well reflect the changes in investor sentiment in the market. Empirical research shows that there is a strong non-linear relationship between investor sentiment and stock returns, with a positive correlation in the short term and a negative correlation in the long term. Using data from China's A-share market from 2010 to 2023, it was found that the VAR prediction model with the sentiment index was 15%-20% better than the traditional financial model in terms of yield prediction, especially when the market fluctuated greatly.

The significance of this study lies in enriching the research on the impact of investor sentiment in behavioral finance and providing strong support for the theory of noise trading. From an application perspective, it can provide a reference for investors' investment decisions, risk management, and the formulation of relevant policies by the government. However, there are certain shortcomings in this study, such as the need for improvement in the design of sentiment indicators and the need for enhancement in the robustness of the prediction model. Future research can continue to be improved and refined in the following areas: First, find more accurate ways to measure emotions, especially those based on big data and artificial intelligence. Secondly, expand the scope of research to the international market to test whether the influence of emotions is universal. Finally, further explore the process of emotion transmission for different market conditions in order to better establish a more stable predictive model.

References

[1] Maurya, P. K., Bansal, R., & Mishra, A. K. (2026). Investor sentiment and its implication on global financial markets: a systematic review of literature. Qualitative Research in Financial Markets, 18(1).

[2] Sun, X., Shen, Y., Li, L., & Suo, W. (2023). Multidimensional network characteristics of earnings, volatility and investor sentiment: A case study of China's energy listed companies. Systems Science and Mathematics, 43(7), 1713–1729.

[3] Yu, J. (2023). Research on metaverse stock price prediction based on investor sentiment: A case study of Qianfang Technology. Wealth Times, (2), 78–81.

[4] Yin, H., & Yang, Q. (2022). The impact of stock forum investor sentiment on stock price bubbles based on Bi-LSTM model mining. Journal of Management, 19(12), 1874–1885.

[5] Xu, W., Huang, J., Fu, Z., & Zhang, W. (2022). A Black-Litterman portfolio model based on sentiment mining in financial texts: A case study of Eastmoney stock forum posts and China's A-share market. Journal of Operations Research, 26(4), 1–14.

[6] Zhao, C., Liu, J., Wu, M., & Chang, X. (2023). Research on LSTM stock price prediction based on numerical and textual features. Journal of Shanxi University (Natural Science Edition), 46(5), 1058–1063.

[7] Gao, Y., Shen, Y., & Xu, J. (2022). The impact of investor sentiment on the yield of the sci-tech innovation market: A text data mining perspective. Operations Research and Management, 31(2), 184–190.

[8] Chen, N. (2021). Construction of investor sentiment indicators considering stock market volatility and their effectiveness in stock market forecasting [Master’s thesis]. Donghua University.

[9] Wang, D., Tian, S., Zhu, J., & Xu, Y. (2021). Research on the functional construction method of investor sentiment index in Chinese stock market. Journal of Mathematical Statistics and Management, 40(1), 162–174.

[10] Jiang, Y. (2020). Construction and validity verification of investor sentiment index in China's stock market. Market Research, (2), 17–19.

[11] Xu, H., & Zhou, W. (2018). Sentiment index and market returns: An analysis incorporating the China Volatility Index (iVX). Journal of Management Sciences in China, 21(1), 88–96.

[12] Wu, Y., & Han, L. (2007). Incomplete rationality, investor sentiment and the mystery of closed-end funds. Economic Research Journal, (3), 117–129.