

Exploring the Impact of Investor Sentiment on the CSI 300 Index Returns Based on Textual Data

Yuxin Rao

*Xiamen University, Xiamen City, China
1164471545@qq.com*

Abstract: In financial markets, understanding the interaction between investor sentiment and stock performance is crucial for investment and regulatory strategies. This paper focuses on the relationship between investor sentiment and stock returns. It collects data from the Eastmoney Stock Bar from July 31, 2015, to December 31, 2024, and constructs sentiment indices. A Vector Autoregression (VAR) model is then established. Through statistical statistics, it finds that there are co-movement trends between investor sentiment and stock returns. The results show that the lagged terms of investor sentiment affects stock return with different directions and degrees. The impulse response analysis indicates that sentiment shocks have short-term impacts on stock returns, trading volume, and transaction amount, which weaken over time. In conclusion, investor sentiment influences stock returns, but its impact is limited. This study provides decision-making references for investors, financial institutions, and regulatory authorities.

Keywords: Investor sentiment, Stock returns, VAR model

1. Introduction

In the Chinese contemporary capital market, information asymmetry hampers efficiency. A gap exists between firms and investors, who lack knowledge and struggle with valuations, and herd mentality further fuels market irrationality. In the digital era, online platforms enhance information exchange but also heighten investors' emotional swings, with emotional contagion in extreme markets stirring irrational sentiment and abnormal price fluctuations, threatening market stability.

Currently, there is extensive academic research on the link between investor sentiment and stock returns. Early studies were theoretical, but with data acquisition tech progress, empirical research has dominated. The recent boom of internet tech provides a new angle, making investor sentiment and online platform data academic focal points. Jin et al. [1] constructed an indicator reflecting investor interaction to explore its internal relationship with stock returns. Liu[2] pointed out that the investor sentiment factor can accurately predict changes in stock returns. In this research context, instead of relying on traditional limited data acquisition channels, this paper employs web crawler software to collect data in real-time. Additionally, by utilizing sentiment lexicons and semantic analysis models, it can precisely extract investor sentiment features for comprehensive characterization and dynamic monitoring.

Within the context of China's capital market, this paper focuses on the CSI 300 Index. Employing big data analysis, natural language processing, and machine learning, it extracts investor sentiment indicators by mining textual data from Eastmoney Stock Bar. Through constructing a VAR model, it

analyzes the interaction and influence mechanism between investor sentiment and stock returns. This research has two major significances. Firstly, it enriches the understanding of the financial market mechanism, providing a new perspective to explain market irrationalities and bridging financial theory with real market operations. Secondly, its findings aid practical decision-making for different market players. Individual investors can make more rational choices and optimize investments, financial institutions can better forecast market trends and manage risks, and regulators can craft targeted policies to safeguard market stability.

2. Literature Review

Investor sentiment refers to the subjective emotional bias of financial market participants during investment decision-making, encompassing tendencies such as optimism or pessimism. It's not rooted in rational asset fundamental analysis but affected by psychological factors and market sentiment, significantly shaping investment behavior and driving fluctuations in asset prices. Shiller [3] described investor sentiment as a social dynamic phenomenon, believing that investor sentiment (such as optimism or pessimism) can cause stock prices to deviate from fundamental values, thus having a profound impact on market operations.

Significant progress has been achieved in this field. For instance, Lee et al. [4] showed that the closed-end funds' discount rate can serve as a proxy for investor sentiment, with a rising rate indicating low sentiment. However, due to the influence of multiple factors on trading data, a single indicator cannot comprehensively and accurately measure it. Tetlock [5] analyzed sentiment in The Wall Street Journal column articles, finding that more negative words suggest lower market returns. This real-time market sentiment capture method may see result accuracy affected by text semantic complexity and sentiment analysis algorithm limitations. Yang et al. [6] created a local attention index via IP address identification, extracting sentiment tendencies with computer text mining to construct investor sentiment indicators. Yao et al. [7] developed a Chinese sentiment lexicon for formal and informal financial texts through lexicon recombination and deep-learning algorithms, then established investor sentiment indicators for listed company annual reports and social media. Yin and Kou [8] built a daily-updated investor sentiment indicator for sample stocks using the Naive Bayes text classification method, identifying sentiment in Eastmoney Stock Bar posts and examining its impact on stock idiosyncratic risks.

In the financial field, the relationship between investor sentiment and stock returns has long been a research focus. With market evolution and research method innovation, many scholars have studied this. Brown and Cliff [9] used AAI investor sentiment survey data and found a correlation between sentiment indicators and market returns, but survey method issues like subjective arbitrariness and sample bias may distort results. Baker and Wurgler [10] noted that high sentiment overvalues hard-to-price stocks, and low sentiment undervalues them. Bu et al. [11] studied investor sentiment's impact on China's stock market, using Eastmoney Stock Bar data and the Naive Bayes method to create a sentiment indicator combining attention and stock review expectations, analyzing its predictive ability. Xue [12] explored the A-share market relationship, creating a sentiment index with Python and text mining for regression-based empirical analysis.

In conclusion, substantial progress has been made in measuring investor sentiment indicators and studying their impacts. These studies deepen our understanding of financial market irrationalities and offer crucial references for investor decisions, financial institutions' risk management, and regulatory policy-making. The relationship between investor sentiment and stock returns is intricate, influenced by numerous factors and operating through various channels. Despite scholars' achievements, there is room for improvement. Therefore, this study seeks to advance research methods by integrating multiple sentiment lexicons. It aims to enhance the accuracy and reliability of sentiment score

calculations, providing a stronger basis for in-depth analysis of the complex relationship between investor sentiment and stock returns.

3. Research Design

3.1. Data Source

3.1.1. Acquisition and Preprocessing of Text Data

The test analysis data in this paper comes from Eastmoney Stock Bar, a prominent Chinese financial community platform where numerous investors gather and engage in real-time discussions about various financial products. Its data is highly representative and timely, providing a reliable reflection of investor sentiment and viewpoints within the financial market. This paper centers on the CSI 300 Index and collects all relevant posts in its stock bar as research samples, spanning from July 31, 2015, to December 31, 2024. During data crawling, Octopus software was utilized for its efficiency and precision in extracting the required data from web pages. After data acquisition, Python was used to preprocess the text for sentiment analysis, with the specific steps as follows.

Text Segmentation and Part-of-Speech Tagging: The Jieba tool was selected to perform word segmentation and part-of-speech tagging on Chinese texts.

Constructing the Sentiment Lexicon: Financial texts contain valuable emotional and intonation data. It's from listed company managers in decision-making and individual investors' emotions from their insights. These emotions influence stock market movements. Constructing effective indicators requires the development of a specialized financial sentiment dictionary. Jiang et al. [13] studied the text emotions in financial news media reports and constructed a Chinese financial sentiment lexicon using the word2vec algorithm. This lexicon contains 9,228 words, with 5,890 negative words and 3,338 positive words. Yao et al. [7] constructed a Chinese sentiment lexicon specifically for the financial field. They built an informal language sentiment lexicon for comments on social media, which contains 965 negative words and 912 positive words. Due to the insufficient coverage of a single lexicon, this paper combines the lexicons of Jiang et al. and Yao et al. and deletes duplicate words to form the reference sentiment lexicon of this paper.

Calculating Sentiment Scores: Taking the text segmented and tagged by Jieba as input, we apply an algorithm to calculate the sentiment score of each sentiment word. The specific calculation steps are as follows:

First, according to the pre-set emotional attributes in the lexicon, words that clearly express positive emotions are assigned a value of 1, and words that convey negative emotions are assigned a value of -1.

Second, the focus is on the roles of degree adverbs and negative words. Degree adverbs, such as "completely", "thoroughly", "100%", etc., can significantly enhance or weaken the emotional intensity expressed by sentiment words. Negative words, such as "negate", "not", etc., can reverse the emotional tendency. Therefore, for words with negative words, the fundamental score of the vocabulary is multiplied by -1.

Standardizing Sentiment Scores: Finally, the scores of all sentiment words in each comment are weighted and summed to obtain the overall sentiment score of the comment. To enhance the comparability and interpretability of calculated sentiment scores, they are normalized, and the quantity of different types of posts is counted. Post-normalization, a score above zero signifies positive emotion, meaning the text conveys a positive stance towards the discussed financial object. A score below zero indicates negative emotion, revealing negativity in the text. When the score equals zero, it represents a neutral emotion, implying the text shows no clear positive or negative sentiment.

3.1.2. Acquisition of Market Index Data

The market index data used in the research covers key variables such as the daily price change rate, trading volume, and trading amount of individual stocks. The data is obtained from the China Research Data Service Platform (CNRDS), and the time span is consistent with that of the stock bar posts.

3.2. Variable Selection

In the empirical analysis, this paper selects and processes key variables to examine the relationship between sentiment in financial texts and stock returns.

3.2.1. Explanatory Variables

Following Antweiler and Frank's [14] method of constructing a bullish indicator from stock bar post classification, the core idea is to gauge investors' bullishness by counting posts of various emotional types. Specifically, the formula for constructing the stock review bullish indicator is as follows:

$$B_t = \frac{M_t^{pos} - M_t^{neg}}{M_t^{pos} + M_t^{neg}} \quad (1)$$

Here, pos represents positive emotions, neg represents negative emotions, and neu represents neutral emotions. M_t^c represents the total number of messages of the corresponding emotion type in period t. The value of this indicator B_t ranges from -1 to 1. The magnitude of its value reflects the relative bullish degree of investors, and this indicator is independent of the total number of posts, which can reflect investors' bullish expectations for the market to a certain extent. The result of B_t in this paper is named *sentiment1*.

In addition, Antweiler and Frank [14] also defined another indicator:

$$B_t^* = \ln \left[\frac{1 + M_t^{pos}}{1 + M_t^{neg}} \right] \quad (2)$$

And through derivation, it is obtained that $B_t^* \approx B_t \ln(1 + (M_t^{pos} + M_t^{neg}))$. The B_t^* indicator not only considers the relative bullish degree but also takes into account the number of posts expressing bullish and bearish views. Their research shows that the B_t^* indicator performs better in measuring investor sentiment. Therefore, this paper mainly reports the results measured by the B_t^* indicator. The result of B_t^* in this paper is named *sentiment2*.

3.2.2. Explained Variable

The daily price change rate (*percentage_change*) of the CSI 300 Index (000300) is taken as the explained variable in this paper.

3.2.3. Control Variables

The control variables are the daily trading volume (*daily_volume*) and trading amount data (*daily_transaction_amount*) of the CSI 300 Index during the corresponding period, and they are logarithmically processed.

3.3. Vector Autoregression Model

The Vector Autoregression (VAR) model is a statistical model that extends the univariate autoregression to a multivariate time series by regarding each endogenous variable as a function of

the lagged values of all endogenous variables. The general form of a p-order VAR model can be expressed as:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t \quad (3)$$

where Y_t is a $k \times 1$ vector of endogenous variables, t represents time; $A_1, A_2, \dots, A_p$ are $n \times n$ lag coefficient matrices, which are used to describe the influence of endogenous variables of different lag orders on the current variable; ε_t is an $n \times 1$ vector of random disturbance terms with a mean of zero.

To deeply analyze the mechanism of the impact of investor sentiment on stock returns, this study takes the return rate of the CSI 300 Index as the explained variable (denoted as *percentage_change*), the sentiment index 2 generated based on the sentiment lexicon as the explanatory variable (denoted as *sentiment2*), and selects the trading volume (*daily_volume*) and trading amount (*daily_transaction_amount*) as control variables to construct VAR model to explore the impact of sentiment on stock returns.

4. Statistical Data and Regression Results

4.1. Descriptive Statistics

To reflect the data of the recent market situation, this paper selects the relevant data from July 1, 2024, to December 31, 2024. By using line charts, it explores the relationships among the price change rate, trading volume, and sentiment index, visually presenting the correlations among the variables.

Figure 1 reveals a strong link between *sentiment1* and the *percentage_change*, reinforcing the inherent relationship between investor sentiment and the price change rate. For instance, from July to September 2024, both *sentiment1* and the *percentage_change* fluctuated multiple times, however, their change directions were fairly consistent in general. Figure 2 illustrates the co-movement between *sentiment1* and *daily_volume*. Since late August, as *sentiment1* rose steadily, *daily_volume* also trended upward steadily until early September. This suggests that high market sentiment may boost investors' trading enthusiasm, thereby increasing trading volume.

Figure 3 indicates that from July to December 2024, the change trends of *sentiment2* and the *percentage_change* were closely aligned. For example, in mid-October 2024, *sentiment2* had significant fluctuations, and the *percentage_change* fluctuated with a comparable amplitude simultaneously. Figure 4 shows that between July and December 2024, *sentiment2* and *daily_volume* had a certain co-movement trend. Notably, in late September 2024, an increase in *sentiment2* was accompanied by a rise in *daily_volume*.

In summary, descriptive statistics of data from July 1 to December 31, 2024, and line chart analysis reveal clear co-movement trends between the price change rate, trading volume, and both *sentiment1* and *sentiment2*. These charts provide an intuitive visual basis for subsequent regression analysis and preliminarily indicate the sentiment index's potential impact on financial market variables. Since *sentiment1* and *sentiment2* perform similarly, to simplify the analysis and highlight key points, the following research will focus on *sentiment2* to elaborate the research process.

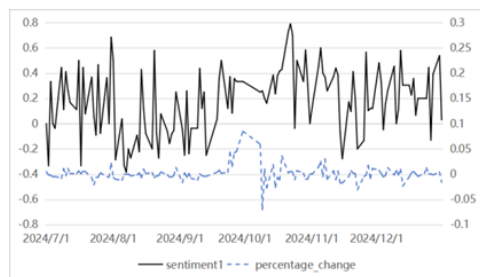


Figure 1: sentiment1 and percentage_change

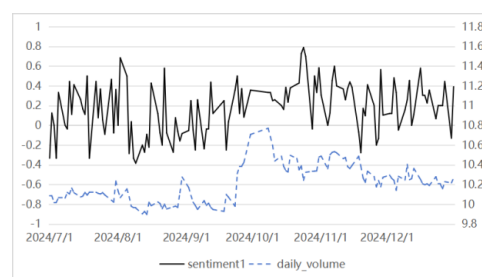


Figure 2: sentiment1 and daily_volume

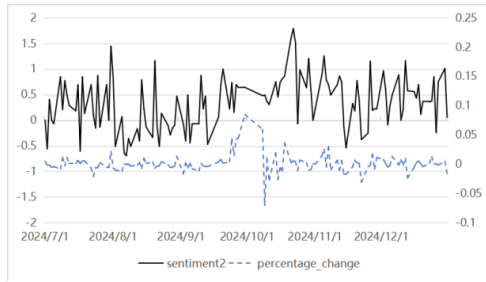


Figure 3: sentiment2 and percentage_change

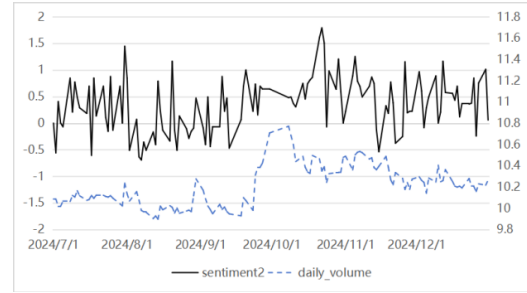


Figure 4: sentiment2 and daily_volume

4.2. Construction of VAR Model

4.2.1. Unit Root Test

Before constructing VAR model, conducting a unit root test on time-series data is a crucial step. This step is to ensure the stationary of the data and prevent the problem of spurious regression. This paper uses the ADF test method to conduct unit root tests on each variable, and the test results are shown in the following Table 1.

Table 1: ADF Test Results

Variable	ADF	P-value	Stationarity
percentage_change	-46.24847	0.0000	Stationary
sentiment1	-21.19653	0.0000	Stationary
sentiment2	-15.13111	0.0000	Stationary
daily_volume	-7.083088	0.0000	Stationary
daily_transaction_amount	-5.781917	0.0000	Stationary

Judging from the ADF test results in Table 1, the ADF values of all variables are significantly smaller than the critical values at different significance levels, and the P-values are all 0.0000, rejecting the null hypothesis. The test results show that the time-series data of the above variables have stationary characteristics. Therefore, VAR model can be constructed to effectively analyze the dynamic relationships among the variables.

4.2.2. Determining the Optimal Lag Order

In the process of constructing VAR model, determining the optimal lag order is a crucial step. This paper employs multiple information criteria to find the VAR model's optimal lag order, with the specific results presented in Table 2. Thus, a lag of 3 orders is chosen as the model's lag order in this paper. As the unit root test reveals that all variables are stationary time-series, satisfying the VAR model's construction conditions, cointegration testing is unnecessary.

Table 2: Determination of the Lag Order

Lag	LogL	LR	FPE	AIC	SC	HQ
0	7167.005	NA	2.09e-08	-6.330539	-6.320421	-6.326847
1	12019.16	9682.861	2.91e -10	-10.60465	-10.55405	-10.58619
2	12215.43	390.9789	2.49e -10	-10.76397	-10.67290	-10.73074
3	12300.01	168.1907	2.34e -10	-10.82458	-10.69304*	-10.77658
4	12342.39	84.12241	2.29e -10	-10.84789	-10.67588	-10.78512*

Table 2: (continued).

5	12371.24	57.16587	2.26e -10	-10.85925	-10.64676	-10.78171
6	12396.69	50.34552	2.24e -10	-10.86760	-10.61464	-10.77530
7	12412.59	31.37871	2.24e -10	-10.86751	-10.57408	-10.76044
8	12432.15	38.55521*	2.23e -10*	-10.87066*	-10.53675	-10.74882

4.2.3. Construction of VAR Model

Taking the price change rate of the CSI 300 as the explained variable (*percentage_change*), the sentiment index based on the sentiment lexicon as the explanatory variable (*sentiment2*), and the trading volume (*daily_volume*) and trading amount (*daily_transaction_amount*) as control variables, the model results are shown in Table 3.

Table 3: Vector Autoregression Model Results

Variable	Coefficient
<i>sentiment2(-1)</i>	0.000558
<i>sentiment2(-2)</i>	-0.000277
<i>sentiment2(-3)</i>	0.000520
<i>percentage_change(-1)</i>	0.016440
<i>percentage_change(-2)</i>	0.027213
<i>percentage_change(-3)</i>	-0.006228
<i>daily_volume(-1)</i>	-0.009650
<i>daily_volume(-2)</i>	0.012797
<i>daily_volume(-3)</i>	-0.001399
<i>daily_transaction_amount(-1)</i>	0.010809
<i>daily_transaction_amount(-2)</i>	-0.010107
<i>daily_transaction_amount(-3)</i>	-0.004283
Constant	0.022758

The research results indicate that the lagged terms of the sentiment index, the price change rate of the CSI 300 itself, and the trading volume and trading amount at different lag orders all impact the price change rate of the CSI 300. These impacts vary in direction and degree. Some lagged terms have a positive impact, while others have a negative impact.

4.2.4. Unit Circle Test

In this paper, a unit circle test is conducted on the constructed VAR model, and the test results are shown in the following figure 5. If all eigenvalues are inside the unit circle, it indicates that the VAR model is stable. If any eigenvalue is outside the unit circle, the model is unstable. As can be seen from the figure, all the characteristic roots are within the unit circle, and the value corresponding to the outermost point is 0.990152, which is less than 1. This result shows that the VAR model constructed in this paper meets the stability conditions and is a stable model.

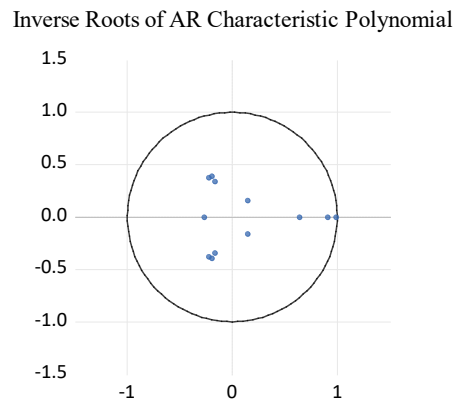


Figure 5: Unit Root Diagram

4.2.5. Granger Causality Test

The Granger causality test is a commonly used method to determine whether there is a causal relationship between two variables. Its core idea is that if the lagged terms of one variable can significantly affect another variable, then the former variable is the Granger cause of the latter variable. In this paper, based on the constructed VAR model, a Granger causality test is conducted on the relationship between the return rate and investor sentiment. The specific test results are shown in following Table 4.

Table 4: Granger Causality Test Results

Hypothesis	Lag Order	F-statistic	P-value	Conclusion
The return rate is not the Granger cause of investor sentiment	3	21.51318	0.0001	Reject the null hypothesis
Investor sentiment is not the Granger cause of the return rate	3	4.457870	0.2161	Accept the null hypothesis

The Granger causality test results show that when the lag order is 3, there is a significant Granger causal relationship from the return rate to investor sentiment ($F = 21.51318$, $P = 0.0001$), but the impact of investor sentiment on the return rate is not significant ($F = 4.457870$, $P = 0.2161$). The return rate is the Granger cause of investor sentiment, while investor sentiment isn't of the return rate. This one-way causality aligns with Wang's [15] finding, where the return rate can predict sentiment changes, but not vice versa. However, this one-way link may be limited by methodology. The standard Granger causality test uses a fixed-parameter VAR model, assuming constant variable relations in the sample period. In reality, sentiment's effect on returns can vary over time, depending on market conditions or external factors. For example, sentiment may only affect returns currently, or be swayed by macro-environment and market policy structures, leading to dynamic adjustments in variable relations.

4.2.6. Impulse Response Diagram

In a VAR model system, the impulse response function describes the dynamic effects on other variables when one variable in the system is shocked by one standard deviation. The following figure shows the responses of the *percentage_change*, *daily_transaction_amount* and *daily_volume* under the shock of one standard deviation of *sentiment2*, with a confidence interval of 95%.

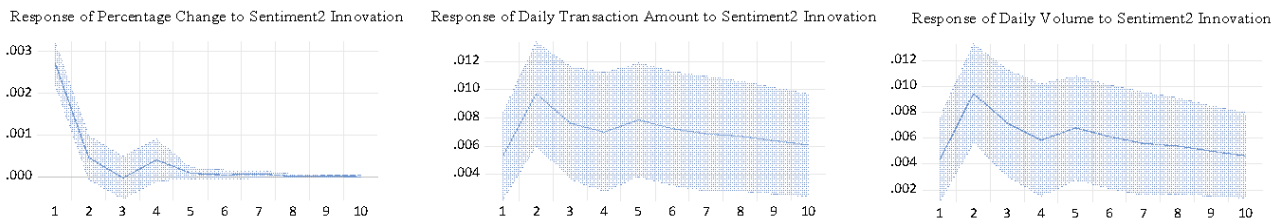


Figure 6: Response to Cholesky One S.D. (d.f. adjusted) Innovations 95% CI using analytic asymptotic S.E.s

As shown in the figure 6, a one-standard-deviation shock to the sentiment index elicits a significant positive response in the *percentage_change* in the first period, followed by a rapid decline. It stabilizes from the second to the third period, then fluctuates minimally and approaches 0. The *daily_transaction_amount* shows an obvious positive response in the first period, peaks in the second, and then gradually declines, trending downward with fluctuations. Similarly, the *daily_volume* has an obvious positive first-period response to the shock, reaches its maximum in the second period, then gradually decreases and stabilizes.

Overall, the shock of the sentiment index has different degrees of short-term impacts on the *percentage_change*, *daily_transaction_amount* and *daily_volume*, but these impacts will gradually weaken over time.

4.3. The Impact of Investor Sentiment on Stock Price Changes

Based on the research above, investor sentiment impacts stock price changes, but isn't decisive. In the short run, sentiment changes can cause stock price fluctuations, but this effect is usually short-lived and of limited magnitude. Also, stock returns can influence investor sentiment, indicating market performance may be key to sentiment swings. For investors, their emotional state can serve as an investment decision reference, but over-reliance is unwise. When making decisions, they should also consider various other factors like a company's fundamentals and the macro-economic development trend.

5. Conclusions and Suggestions

This study delved into the relationship between investor sentiment and stock price changes in China's capital market. Analyzing data from July 31, 2015, to December 31, 2024, through descriptive statistical analysis and line charts, it found obvious co-movement trends between the stock return and sentiment index, offering preliminary and intuitive evidence for further research.

After constructing a VAR model and running a series of tests, it was found that the lagged sentiment index, CSI 300 price change rate, and trading volume/amount at different lags all affected the CSI 300 price change rate, with varying directions and degrees. The impulse response diagram results showed that a shock to the sentiment index had different short-term impacts on the *percentage_change*, *daily_transaction_amount* and *daily_volume*, which gradually weakened over time. This suggests that investor sentiment fluctuations cause short-term financial market trading indicator fluctuations, but the impact fades in the long run.

Based on the above research findings, to foster the healthy and stable growth of the financial market and guide investors to invest rationally, the following suggestions are proposed:

(1) When investing, investors should refrain from basing decisions solely on market sentiment. While it can influence stock prices in the short term, it's not decisive. They should focus more on in-depth analysis of a company's financials, profitability, and market competitiveness. Also, they

need to track macro-economic changes. By considering these factors together, investors can make more scientific and rational decisions, reduce risks, and increase the odds of investment success.

(2) Financial institutions should enhance market sentiment monitoring and analysis. By leveraging advanced technology and data analysis to grasp market sentiment trends, they can allocate assets, optimize portfolios, and mitigate risks. Moreover, they can offer investors more personalized investment consulting services based on market sentiment fluctuations, aiding in more rational investment decisions.

(3) Regulatory authorities should highly value market sentiment's impact on financial market stability. Build and improve the market monitoring system to track sentiment changes in real-time. Based on the monitoring results, formulate effective policies to ensure the market's stable and healthy development. Also, strengthen supervision and severely punish illegal acts like market manipulation and insider trading for a fair and transparent market.

References

- [1] Jin, D. H., & Li, Y. *Investor Interaction and Stock Returns: Empirical Evidence from Social Media* [J]. *Financial Forum*, 2017, 22(05): 72-80. DOI: 10.16529/j.cnki.11-4613/f.2017.05.008.
- [2] Liu, Q. *The Impact of Investor Sentiment on Stock Returns: An Analysis of Stock Bar Comments* [J]. *Science Technology and Industry*, 2024, 24(24): 203-209.
- [3] Shiller, R. J. *Stock prices and social dynamics*. *Brookings Papers on Economic Activity*, 1984(2), 457-498.
- [4] Lee C, Shleifer A, Thaler R H. *Investor sentiment and the closed-end fund puzzle* [J]. *Journal of Finance*, 1991, 46(1): 75 -109.
- [5] Tetlock P C. *Giving Content to Investor Sentiment: The Role of Media in the Stock* [J]. *Journal of Finance*, 2007, 62(3): 1139 -1168.
- [6] Yang, X. L., Shen, H. B., & Zhu, Y. *Local Bias, Investor Sentiment and Stock Returns: Empirical Evidence from Internet Forums* [J]. *Journal of Financial Research*, 2016, (12): 143 -158.
- [7] Yao, J. Q., Feng, X., Wang, Z. J., et al. *Intonation, Emotion and Market Impact: Based on Financial Sentiment Lexicon* [J]. *Journal of Management Sciences in China*, 2021, 24(05): 26-46.
- [8] Yin, H. Y., & Kou, W. J. *Measurement of Investor Sentiment Based on Naive Bayes Method and Its Impact on Stock Idiosyncratic Risk* [J]. *Chinese Journal of Management Science*, 2024, 32(04): 38-47.
- [9] Brown G W, Cliff M T. *Investor sentiment and the near-term stock market* [J]. *Journal of Empirical Finance*, 2004, 11(1): 1-27.
- [10] Baker M, Wurgler J. *Investor sentiment and the cross-section of stock returns* [J]. *Journal of Finance*, 2006, 61(4): 1645 -1680.
- [11] Bu, H., Xie, Z., Li, J. H., et al. *The Impact of Investor Sentiment Based on Stock Reviews on the Stock Market* [J]. *Journal of Management Sciences in China*, 2018, 21(04): 86 -101.
- [12] Xue, B. Q. *Analysis of the Relationship between Stock Market Returns and Investor Sentiment in China Based on Text Mining Method* [J]. *China Management Informationization*, 2024, 27(20): 118 -121.
- [13] Jiang, F. W., Meng, L. C., & Tang, G. H. (2021). *Media Text Sentiment and Stock Return Prediction*. *China Economic Quarterly*, 21(04), 1323-1344. doi:10.13821/j.cnki.ceq.2021.04.10.
- [14] Antweiler W, Frank M Z. *Is all that talk just noise? The information content of Internet stock message boards* [J]. *Journal of Finance*, 2004, 59(3): 1259 -1294.
- [15] Wang Y H, Keswani A, Taylor S J. *The relationships between sentiment, returns and volatility* [J]. *International Journal of Forecasting*, 2006, 22(1), 109-123.