Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/88/20241662

A study on the impact of investor sentiment on individual stocks based on text mining

Jiamei Li

China University of Geosciences (Beijing), No. 29 Xueyuan Road, Haidian District, Beijing

594765550@qq.com

Abstract. This paper investigates the influence of investor sentiment on individual stock prices through text mining. By reviewing relevant theories and literature, we utilized web crawler technology to collect comment data from the Eastmoney Stock Bar, constructing a comprehensive index of investor sentiment and an index of investor sentiment for individual stocks. Employing the VAR model and Granger causality test, we analyzed the correlation and dynamic relationship between investor sentiment and stock price fluctuations. The study found that stocks of different types of financial institutions react differently to investor sentiment, and that investor sentiment significantly affects stock prices. This paper puts forward corresponding policy recommendations, which can help market investors make scientific decisions and promote the healthy development of the market.

Keywords: Stock Market, Investor Sentiment, Text Mining, VAR Model.

1. Introduction

1.1. Research Background and Current Status

The stock market is a special type of commodity market, with stocks as the main trading objects. It reflects the operational status of enterprises and market expectations through changes in supply and demand relationships. The scale of the stock market continues to expand, the market system is increasingly perfected, market functions become more apparent, and the market ecosystem is continuously optimized, providing strong support for the transformation and high-quality development of the national economy. As of October 31, 2023, the total number of listed companies in China has exceeded 5,300, with a total market value close to 80 trillion yuan, ranking second in the world. In modern information society, the internet has become the primary way to obtain information. The China Internet Network Information Center (CNNIC) recently announced the latest "Statistical Report on the Development of the Internet in China" in Beijing, numbered the 52nd issue. The report shows that by June 2023, the total number of Chinese internet users had risen to 1.079 billion, an increase of 11.09 million users compared to December 2022. The views and comments on stocks posted by internet users in financial forums related to stocks and securities can directly reflect the emotions and behaviors of investors and have an important impact on the trend of the stock market. The popularity and development of the internet have enabled investors to obtain and share financial information on online platforms. This information not only contains a wide range of financial dynamics but also deeply reflects the fluctuations

DOI: 10.54254/2755-2721/88/20241662

and changes in investors' emotions, which have a significant impact on market participants. It can guide investors to make wiser decisions and affect the overall market trend. Currently, many studies have shown a close relationship and mutual influence between financial information on the internet and stock market fluctuations. Based on this, this paper takes the stock market as the research object, the impact of investor sentiment on the stock market as the research question, and text mining as the starting point, to mine text information from stock forum netizen comments, deduce their emotional tendencies, and the correlation with stock price fluctuations, to assist market investors in making scientific decisions, reducing risks, and promoting the healthy development of the market, which has innovation and practical significance.

1.2. Research Content

This study aims to explore the mechanism and effect of investor sentiment on the Chinese ChiNext market.

Firstly, this paper reviews the relevant theories of text mining and investor sentiment, analyzes the concept of investor sentiment, measurement methods, and the impact path on stock market fluctuations in previous studies, and extracts relevant information that is helpful for the research content of this paper for reference, so as to construct a theoretically supported research framework and improve the theoretical foundation of this paper.

Secondly, this paper uses web crawler technology to collect comments from stock friends in the stock bar, screens and labels the main topic tags of stock reviews; uses text mining technology to extract emotional tendencies and intensity from comments, constructs a comprehensive index of investor sentiment in the stock market and an index of investor sentiment for individual stocks. The VAR model is used to empirically test the degree and direction of the impact of investor sentiment on the stock market, as well as the dynamic relationship between investor sentiment and the stock market.

Lastly, by using regression and correlation analysis, the sensitivity of individual stocks of different types of financial institutions to investor sentiment is compared and the reasons are analyzed, the impact of investor sentiment on stock price fluctuations is proposed, and corresponding policy recommendations and research prospects are put forward on this basis.

2. Literature Review

The importance of macroeconomic conditions for the operation of the stock market has been well documented in the literature [1]. For example, Ahmed revealed that there is a long-term relationship between stock market indices and macroeconomic variables such as industrial production, exports, foreign direct investment, money supply, exchange rates, and interest rates [2]. Wongbangpo and Sharma found that stock price indices are positively correlated with macroeconomic variables in the long run [3]. Stock prices often respond to changes in macroeconomic factors. Nasseh and Strauss found that the major stock markets surveyed are highly influenced by domestic and international macroeconomic activities, such as industrial production, business surveys of manufacturing activities, short-term and long-term interest rates, and foreign macroeconomic activities [4]. Stock prices and output. Therefore, knowledge about the relationship between macroeconomic factors and the operation of the stock market provides opportunities to earn substantial returns by selecting specific stocks when information on certain macroeconomic variables is available.

In the history of the financial market, there have been many phenomena of sharp rises, sharp declines, and crashes in the stock market, which cannot be justified by the macroeconomic level alone. Early on, Keynes conducted a preliminary analysis of speculative markets and investor sentiment [5]. Scholars such as Murphy also recognized the importance of investor sentiment and established intuitive measurement standards for extreme optimism and pessimism long before modern academic research [6]. Theoretically, DeLong and others have developed various models to understand how investor sentiment is formed and affects the market [7]. Investor sentiment cannot be directly observed and must be estimated using certain technologies and models. Among the existing measurement standards, the discount on closed-end funds proposed by Zweig and Shleifer and others is the oldest measurement

DOI: 10.54254/2755-2721/88/20241662

method [8]. Xu Qin compared subjective and objective measurement indicators as well as comprehensive measurement indicators and other sentiment measurement methods [9].

With the rapid development of the economy, traditional investor sentiment measurement standards can no longer accurately measure investor sentiment, and their predictive ability for the overall stock market is not significant. To this end, Huang and others emphasized the importance of keeping the index consistent with its purpose, and showed that alternative indices using the same proxies as traditional investor sentiment measurement standards can make negative predictions about the stock market, and now higher sentiment indicates lower future market returns [10]. With the rapid development of the Internet, online forums have become a huge collection of valuable opinions and comments, and more and more researchers are taking a keen interest in this issue [11]. In the research conducted by Subrahmanian and Reforgiato, emotional analysis was performed for the first time using adjectives, verbs, and other components [12]; in the research conducted by Shi and others, emotional analysis was performed on different ideas in Chinese online forums, and an algorithm different from the probability word list of classification algorithms such as support vector machines (SVM) and naive Bayes was proposed, which has been proven to have better performance in experiments [13].

Existing text mining and sentiment analysis methods are extremely rich, and scientists' models for dealing with textual emotional tendencies are gradually improving, but there are still some shortcomings. First, there is currently no sentiment dictionary specifically built for stock comments, most of them are sentiment dictionaries for news comments or financial data. Second, the current stock sentiment analysis only considers the impact of the emotional tendencies of investors' comments on stock prices, and does not consider the impact of other factors affecting stock prices, such as closing prices, on stock price fluctuations. To make up for these shortcomings, this study uses principal component analysis based on text mining to comprehensively consider other factors affecting stock prices, extract principal components, calculate relevance and the degree of impact, construct a comprehensive index of investor sentiment and individual stock indices, and use the VAR model, Granger causality test, etc. to test the accuracy of the conclusions, effectively determine the correlation between investor sentiment and stock prices.

3. Establishment of the Investor Sentiment Index for Individual Stocks

3.1. Selection of Individual Stock Data

This paper selects China Construction Bank, Ping An Insurance Company of China, and China Merchants Securities as research subjects of three different types of financial institutions to verify the impact of the investor sentiment index on stock price fluctuations. Banks, securities, and insurance companies are the three major financial institutions that currently attract high public attention, encompassing the vast majority of market investors. Using them as research subjects is representative and provides a good reflection of the sentiment of stock market investors, making the conclusions more persuasive. This paper sets the research period for individual stocks from June 30, 2020, to June 30, 2023, a total of 730 trading days, excluding posts with blank titles and those with non-standard dates.

3.2. Data Extraction

Using Python web crawler compilation, this paper extracts stock friend comment data for the aforementioned three companies' stocks from the Eastmoney Stock Bar forum and processes the data for cleaning, retaining several attributes including author, post time, main post title, views, and comments. Taking China Construction Bank as an example, the final number of valid individual stock text comments retained for China Construction Bank is 39,156. Table 1 displays some stock review data and post titles for China Construction Bank.

DOI: 10.54254/2755-2721/88/20241662

Table 1. China Construction Bank Stock Review Data

Serial No.	Views	Post Title	Author	Date
1	106	"Vibrant Youth, 'Sound' Arrives" — Construction Bank Guangdong Province Jiangmen Branch Holds "Love ·" Time Music Concert	Financial Management	2023-06-30
2	1177	Moutai Liquor Actually Appreciates Little	z19660919qf	2023-06-30
3	3213	Mid-Year Review of Southbound Funds: Accumulated Inflow of 123.5 Billion, Down 40% Year-on-Year; Public Utilities Rebound as Key Sector	Tencent Holdings	2023-06-30
4	305	Jianxin Consumer Finance Company Opens in Beijing, Will Launch Digital RMB Consumer Loan Business	Bank Consulting	2023-06-30

3.3. Index Construction

After obtaining the stock comments for these three stocks during the aforementioned period, the Python Chinese processor SnowNLP library is utilized to process the data. Each post's sentiment score is calculated to reflect investors' emotional tendencies towards the bank. The average sentiment score for each of these six banks on each trading day is computed to obtain the sentiment score average for the stock on the current working day. The processed results yield the sentiment indices for these three stocks, denoted as sent_j, sent_a, and sent_z, respectively.

4. Empirical Analysis of the Impact of Individual Stock Indices on Individual Stocks

4.1. Descriptive Statistics of Data

Table 2. Variable Description

Variable	Meaning
sent_j	China Construction Bank Investor Sentiment Index for Individual Stock
change_j	Change in the Closing Price of China Construction Bank Stock
sent_a	Ping An Insurance of China Investor Sentiment Index for Individual Stock
change_a	Change in the Closing Price of Ping An Insurance of China Stock
sent_z	China Merchants Securities Investor Sentiment Index for Individual Stock
change_z	Change in the Closing Price of China Merchants Securities Stock

This chapter will conduct a study on the impact of the investor sentiment index on individual stocks, with the relevant variables as shown in Table 2, and compare the extent of reaction to investor sentiment across stocks of different financial institutions. Descriptive statistical analysis will be performed on the change and sent variables of these three stocks (i.e., change_j, sent_j, change_z, sent_z, change_a, sent_a), and the results of the statistical analysis are presented in Table 3.

Table 3. Descriptive Statistics of Individual Stocks for the Six Major Banks

	change_j	sent_j	change_z	sent_z	change_a	sent_a
Mean	6.95E-05	0.455458	-0.00043	0.437711211	-0.00042	0.431405
Standard Error	0.00045	0.003047	0.000799	0.002567807	0.000678	0.001708
Standard Deviation	0.012169	0.08233	0.021477	0.069044891	0.018237	0.045921

DOI: 10.54254/2755-2721/88/20241662

Table 3. (continued).

Variance	0.000148	0.006778	0.000461	0.004767197	0.000333	0.002109
Minimum Value	-0.08647	0.206227	-0.18468	0.20564576	-0.06898	0.292033
Maximum Value	0.085329	0.836823	0.100228	0.71968516	0.100022	0.590141

As illustrated, the average values of the investor sentiment indices for individual stocks of the three companies are all close to the critical value of 0.5, indicating that during this period, the distribution of stock investors' sentiment is relatively even and tends to be negative. Moreover, the standard deviation of the closing price changes, change_j, change_z, and change_a, is all less than 0.1%, which preliminarily indicates that these three variables are relatively stable.

4.2. Data Stability and Causality Test

4.2.1. ADF Test

An Augmented Dickey-Fuller (ADF) test is conducted to ensure that the original data set sequences are stationary. This is essential to ensure that the constructed VAR model is meaningful and effective. When the p-value of the ADF test is less than 0.05, the data is considered stationary, allowing for the subsequent construction of the VAR model. In this section, Eviews is used to perform the ADF test, and the results are presented in Table 4. It can be observed that the p-values for the six aforementioned variables are all less than the critical value of 0.05, indicating that these six variables are stationary during the period of this study, and thus a VAR model can be constructed on this basis.

Table 4. ADF Test Results for Individual Stocks

Series	t-Stat	Prob.
CHANGE_J	-25.466	0.0000
SENT_J	-5.3495	0.0000
CHANGE_Z	-24.770	0.0000
SENT_Z	-13.693	0.0000
CHANGE_A	-27.425	0.0000
SENT_A	-4.4458	0.0003

After establishing the stationarity of two variables, this chapter will first determine the lag order of the model before constructing the VAR (Vector Autoregression) model. Taking the change in the closing price of China Construction Bank, denoted as change_j, and its corresponding investor sentiment index, sent_j, as an example, and setting the maximum lag order to 8 days, the results are presented as shown in Table 5.

Table 5. Determination of VAR Lag Order for China Construction Bank Individual Stock

Lag	LogL	LR	FPE	AIC	SC	HQ
0	2978.245	NA	8.60e-07	-8.290375	-8.277627	-8.285453
1	3015.989	75.17441	7.83e-07	-8.384372	-8.346128	-8.369606
2	3036.277	40.29305	7.48e-07	-8.429742	-8.366003	-8.405132
3	3048.708	24.61923	7.31e-07	-8.453226	-8.363991	-8.418772
4	3063.405	29.02457	7.09e-07	-8.483021	-8.368291*	-8.438723*
5	3070.074	13.13500	7.04e-07	-8.490458	-8.350232	-8.436316
6	3077.182	13.95884	6.98e-07*	-8.499115*	-8.333394	-8.435129
7	3079.315	4.176389	7.02e-07	-8.493914	-8.302697	-8.420084
8	3081.157	3.596101	7.06e-07	-8.487902	-8.271189	-8.404228

DOI: 10.54254/2755-2721/88/20241662

In the selection of the optimal lag order for China Construction Bank, both the SC (Schwarz Criterion) and HQ (Hannan-Quinn Criterion) identified a maximum lag order of 4 for the model. These two criteria are generally more reliable, hence this paper determines the lag order of the VAR model for the investor sentiment index of China Construction Bank to be 4. Similarly, the optimal lag orders for the VAR models of China Merchants Securities and Ping An Insurance are both determined to be 2.

4.2.2. Granger Causality Test

The Granger causality test is a commonly used statistical method that is typically employed to detect whether there is a causal relationship between two sets of data. During the test, the optimal lag order is set, and regression analysis is conducted. The test is based on the F-statistic of the model. Utilizing Eviews software for the operation, the results are presented in the following table.

Table 6. Granger Causality Test Results

Dependent variable: CHANGE_J				Dependent variable: SENT_J			
Excluded	Chi-sq	df	Prob.	Excluded	Chi-sq	df	Prob.
SENT_J	5.901263	4	0.0001	CHANGE_J	11.84907	4	0.0185
Dependent variable: CHANGE_Z				Dependent variable: SENT_Z			
SENT_Z	8.20844	2	0.0001	CHANGE_Z	5.657118	2	0.0006
Dependent variable: CHANGE_A				Dependent variable: SENT_A			
SENT_A	1.839193	2	0.0001	CHANGE_A	4.736751	2	0.0136

The results indicate that the p-values for the three banks are all less than 0.05, demonstrating that the closing price changes of the stocks of these three financial institutions and investor sentiment influence each other, exhibiting a reciprocal causal relationship. That is, stock price fluctuations can affect and lead to changes in investor sentiment, and conversely, investor sentiment can also guide the trend of stock price movements. When the stock market is performing well, investor sentiment tends to be more positive. This positive sentiment encourages them to invest more funds into the stock market, which in turn promotes the stock market's development in a better direction, resulting in higher stock prices.

4.3. Model Stability Test

4.3.1. VAR Model Result Analysis

Taking the VAR model constructed for China Construction Bank as an example and using Eviews software to build a VAR(4) model, the results are as shown in Table 7. Analysis of the results in the table shows an R-value of 20.91%, indicating that the investor sentiment index (sent_j) and the change in closing price (change_j) of China Construction Bank in the previous four periods account for 20.91% of the influence on the current change in investor sentiment. Such a result indicates a significant impact.

Table 7. China Construction Bank VAR(4) Model Results

	CHANGE_Y	SENT_Y
CHANGE_Y(-1)	0.044163	0.412644
CHANGE_Y(-2)	0.026487	0.018547
CHANGE_Y(-3)	-0.023962	-0.595453
CHANGE_Y(-4)	-0.073231	-0.289372
SENT_Y(-1)	0.001725	0.165631
SENT_Y(-2)	0.010493	0.162977
SENT_Y(-3)	-0.003140	0.121022
SENT_Y(-4)	0.006133	0.201586
R-squared	0.016785	0.209195

DOI: 10.54254/2755-2721/88/20241662

4.3.2. VAR Model Stability Test

Utilizing the AR (Autoregressive) root diagram analysis to verify the stability of the VAR model is an essential step in ensuring the validity of the model. The fundamental concept of this method is to plot the roots on a coordinate axis; if all AR characteristic roots are within the unit circle, it indicates that the model is stable. In this section, Eviews software is used for experimental operations to conduct the stability test for the three stocks, as shown in Figures 1 to 3 (i.e., the AR root diagrams).

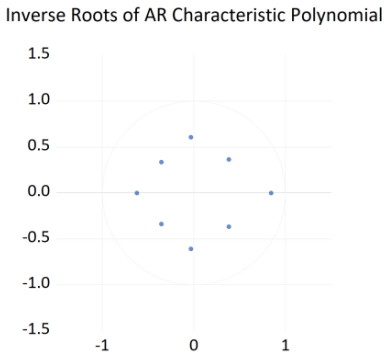

Figure 1. AR Root Diagram of the VAR(4) Model for China Construction Bank

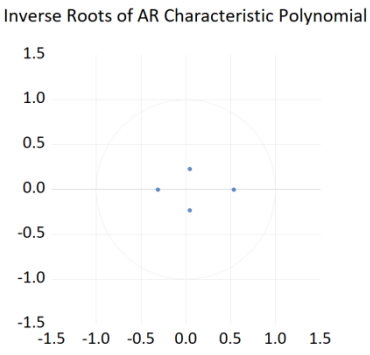

Figure 2. AR Root Diagram of the VAR(2) Model for China Merchants Securities

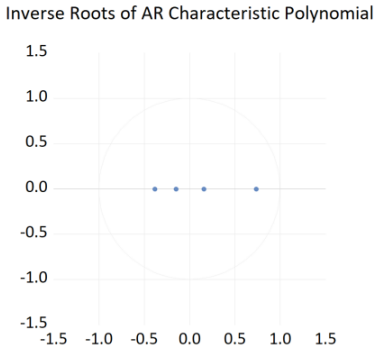

Figure 3. AR Root Diagram of the VAR(2) Model for Ping An Insurance of China

Based on the AR root diagrams mentioned above, it can be clearly observed that all roots of the VAR models for the three companies are within the unit circle required by the AR model, indicating that these three models are stable and reliable, thereby lending greater credibility to the predictions and interpretations based on these models.

4.3.3. VAR Model Cointegration Test

This study employs a cointegration test to verify whether there is an equilibrium relationship between the investor sentiment index for individual stocks and the change in stock closing prices. This analysis not only enhances the understanding of market dynamics but also provides investors with reliable evidence for formulating more precise investment strategies.

Table 8. Cointegration Test of Individual Stock VAR Relationships

Variable	Coefficient	Std. Error	t-Statistic	Prob.	R-squared
CHANGE_Y	1.164858	0.413928	2.814158	0.0050	
C	0.455299	0.005037	90.39780	0.0000	0.016890
CHANGE_Z	0.740399	0.174116	4.252330	0.0000	
C	0.437940	0.003740	117.0877	0.0000	0.041669
CHANGE_A	0.987106	0.171705	5.748850	0.0000	
C	0.431779	0.003127	138.1003	0.0000	0.113950

The test results shown in Table 8 indicate that the p-values obtained from the cointegration test for the stocks held by the three financial institutions are all less than 0.05, leading to the rejection of the null hypothesis that there is no cointegration relationship between the two variables. The existence of this cointegration relationship suggests that in these three companies, the fluctuations in the investor sentiment index for individual stocks will have an impact on the change in closing prices in the long term.

DOI: 10.54254/2755-2721/88/20241662

4.3.4. Impulse Response Analysis

In this study, the PR (Pulse Response) Analysis aims to delve into the interrelationship between the sent-type variables, the investor sentiment index for individual stocks, and the change in daily closing prices. By thoroughly analyzing the impulse response charts, the complex relationship between investor sentiment and individual stocks can be better understood. Figures 4 to 6 display the results of the impulse response analysis.

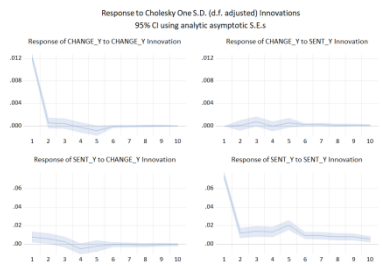

Figure 4. Impulse Response Chart for China Construction Bank

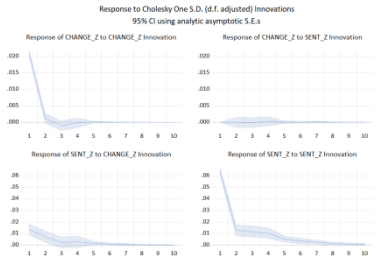

Figure 5. Impulse Response Chart for China Merchants Securities

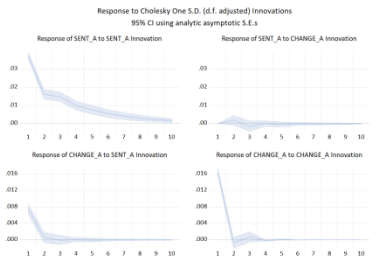

Figure 6. Impulse Response Chart for Ping An Insurance of China

By observing the impulse charts of the model, it can be found that when the sentiment index for individual stocks is subjected to a positive shock, the change in closing price initially shows a strong positive response. However, as time progresses, this positive reaction gradually diminishes until it stabilizes. Conversely, when the closing price is subjected to a positive shock, the sentiment index for individual stocks follows the aforementioned trend. These observations indicate that due to the potential for stock price reversals and other uncertainties in China's stock market at any time, investors should fully consider the impact of sentiment factors on the market when making decisions, and also be vigilant about the risks that may arise from a reversal of market sentiment.

4.4. Analysis of the Differences in the Impact of Sentiment Fluctuations on Financial Institutions

4.4.1. Empirical Study of Regression Analysis

Regression analysis is an important method in statistics used to determine the quantitative relationship of interdependence between two or more variables. This method is commonly used to study the relationship between the dependent variable and independent variables. In this study, we use this method to help understand how investor sentiment affects stock prices, reflecting the extent to which investor sentiment influences individual stocks. The investor sentiment indices for China Construction Bank, China Merchants Securities, and Ping An Insurance, denoted as sent_j, sent_z, and sent_a, are considered as independent variables, while the changes in their stock closing prices, change_j, change_z, and change_a, are considered as dependent variables. The correlations between them are calculated and regression analysis is performed. The results of the regression analysis are shown in Table 9.

Table 9. Regression Analysis Results for Three Types of Financial Institutions

	China Construction Bank	China Merchants Securities	Ping An Insurance
Multiple R	0.535472	0.605497	0.841737
Standard Error	0.081627	0.067618	0.043187

The results in the table show that the correlation coefficient between sent_j and change_j for China Construction Bank is 0.5354, while for China Merchants Securities and Ping An Insurance, the coefficients are 0.6055 and 0.8417, respectively. This indicates that among the three financial industries represented in this study—banking, securities, and insurance—bank stocks are the least sensitive to

DOI: 10.54254/2755-2721/88/20241662

fluctuations in investor sentiment, whereas investor sentiment has a greater impact on the stock prices in the insurance industry.

4.4.2. Regression Result Analysis

I. The stock prices in the insurance industry are most sensitive to changes in investor sentiment. Insurance industry products are closely related to individual lives, and their demand is easily influenced by investors' risk perception and emotions. Optimistic or pessimistic emotions can drive insurance purchase decisions, affecting the performance and stock prices of insurance companies. Insurance is often seen as a risk-avoidance tool, with demand increasing in times of market uncertainty. The structure dominated by individual investors makes insurance investment decisions more emotional. In contrast, the banking and securities industries are greatly influenced by the market, interest rates, and policies, with less impact from investor sentiment. The limited dissemination of insurance information and the exacerbation of information asymmetry lead to more emotional decision-making. In summary, investor sentiment has a direct impact on insurance stock prices, and the limitation of information channels further amplifies this effect.

II. The impact of investor sentiment changes on the stocks held by banks is minimal. Banking operations are stable due to strict regulation, and core businesses such as deposits, loans, settlements, and wealth management are less affected by market fluctuations. The stability of banks is crucial to the economic system, and strict regulation ensures their safety and stability. The products in the insurance and securities industries are diverse and complex, susceptible to market fluctuations and emotional influences. The investor structure in the banking industry is diverse, leading to more rational decision-making; whereas insurance and securities may be more emotional. In terms of market size, the banking industry is vast and has a broad impact, with stock prices being stable and not easily swayed by single emotions; insurance and securities may be more prone to fluctuations.

5. Regression Result Analysis

This paper, based on the research background, current situation of the stock market, and the in-depth analysis of its relationship with investor sentiment, accurately recognizes the importance of the stable operation of the stock market, where investor sentiment is one of the important factors affecting the stock market trend. By using text mining and sentiment analysis techniques, this paper has crawled relevant stock comments from the Eastmoney Stock Bar for China Construction Bank, China Merchants Securities, and Ping An Insurance, and constructed an investor sentiment index for individual stocks through data preprocessing, sentiment analysis, and principal component analysis.

The construction of investor sentiment indices for different types of companies helps to more comprehensively assess the impact of investor sentiment on the stock market. By constructing these indices, this paper can more intuitively understand the trend of investor sentiment changes and predict their impact on the stock market. This is of great reference value for investors to formulate investment strategies and manage risks.

Investor sentiment is one of the important factors affecting the stock market. By deeply analyzing the comments of netizens in the Eastmoney Stock Bar and using the VAR model and Granger causality test techniques, this paper found a significant correlation between the sentiment tendencies of investors and the change in the closing price of individual stocks. When investor sentiment is positive, it often drives the rise of individual stock prices; on the contrary, when investor sentiment is low, individual stock prices may be suppressed. This finding further verifies the important role of investor sentiment in the stock market, where changes in investor sentiment can affect the supply and demand relationship of the market, thereby affecting the trend of stock prices.

This paper also compares the sensitivity of three different types of financial institutions, namely banks, securities, and insurance, to changes in investor sentiment and finds that the stock prices in the insurance industry are most sensitive to changes in investor sentiment, while the impact of investor sentiment changes on the stocks held by banks is minimal. This may be related to the types of investors in the industry, the types of products and business, and market regulation. This result also provides

DOI: 10.54254/2755-2721/88/20241662

decision-making suggestions for investors. In the investment decision-making process, investors should consider the impact of emotional factors and not blindly follow the trend, so as to make more rational and comprehensive judgments.

Finally, there are still some limitations and deficiencies in this study. Due to the limitation of data sources, this paper only selects a limited amount of data to construct indices, which may affect the universality of the research results. Future research can further expand data sources to verify the wide applicability of the research conclusions. In addition, when analyzing the stock review text data, the sentiment processing of Chinese is directly carried out using the SnowNLP library in python, which is mainly constructed for product reviews. Directly using this library may lead to a deviation between the constructed sentiment index and the real sentiment tendency. Future research can try to build a sentiment dictionary specifically for financial stock comments, and use this dictionary for stock review sentiment analysis to obtain more accurate and reliable conclusions.

In summary, this study, through text mining and sentiment analysis techniques, deeply explores the impact of investor sentiment on the stock market and constructs an investor sentiment index for individual stocks. The research results show that investor sentiment is one of the important factors affecting the stock market. Compared with other financial institutions, the insurance industry is more susceptible to the impact of investor sentiment, while banks are relatively stable and less affected by investor sentiment. These findings are of important practical significance for investors to formulate investment strategies and manage risks.

References

[1] Marina Gorbacheva. The Impact of Macroeconomic Factors on the Returns of the Russian Stock Market [D]. Harbin Institute of Technology, 2022.

[2] Zhan Yue. The Impact of Macroeconomic Factors on the Stock Market [J]. Shanxi Agricultural Economy, 2017(14): 36+40.

[3] Zhao Huafeng. Analysis of Macro Factors Affecting China's Stock Market [J]. Oriental Corporate Culture, 2013(19): 93.

[4] Nasseh A, Strauss J. Stock Prices and Domestic and International Macroeconomic Activity: A Cointegration Approach [J]. The Quarterly Review of Economics and Finance, 2000, 40(2): 229-245.

[5] Beckhart B H. The General Theory of Employment, Interest and Money by John Maynard Keynes [J]. Political Science Quarterly, 1936, 51(4): 600-602.

[6] Gifford E. Technical Analysis of the Futures Markets [M]//EVANS G. ICCH Commodities Yearbook 1990. London: Palgrave Macmillan UK, 1990: 23-27.

[7] De Long J B, Shleifer A, Summers L H, et al. Noise Trader Risk in Financial Markets [J]. Journal of Political Economy, 1990, 98(4): 703-738.

[8] Shleifer A, Vishny R W. The Limits of Arbitrage [J]. The Journal of Finance, 1997, 52(1): 35-55.

[9] Xu Qian. A Review of Research on Investor Sentiment Measurement Methods and Influencing Mechanisms [J]. Finance and Accounting Monthly, 2015(9): 87-89.

[10] Huang D, Jiang F, Tu J, et al. Investor Sentiment Aligned: A Powerful Predictor of Stock Returns [M]. Rochester, NY (2014-10-13).

[11] Shi L, Sun B, Kong L, et al. Web Forum Sentiment Analysis Based on Topics [C]//2009 Ninth IEEE International Conference on Computer and Information Technology: Volume 2. 2009: 148-153.

[12] Subrahmanian V S, Reforgiato D. AVA: Adjective-Verb-Adverb Combinations for Sentiment Analysis [J]. IEEE Intelligent Systems, 2008, 23(4): 43-50.

[13] Shi L, Sun B, Kong L, et al. Web Forum Sentiment Analysis Based on Topics [C]//2009 Ninth IEEE International Conference on Computer and Information Technology: Volume 2. 2009: 148-153.