Modern Economy, 2023, 14, 778-795
https://www.scirp.org/journal/me
ISSN Online: 2152-7261
ISSN Print: 2152-7245

Study Based on SNOWNLP Model Mining of Stock Bar Investors' Emotions on Stock Prices

Chenyu Li

College of Management and Economics of Tianjin University, Tianjin, China
Email: chenyutju@gmail.com

How to cite this paper: Li, C. Y. (2023). Study Based on SNOWNLP Model Mining of Stock Bar Investors' Emotions on Stock Prices. *Modern Economy, 14,* 778-795.
https://doi.org/10.4236/me.2023.146042

Received: April 25, 2023
Accepted: June 27, 2023
Published: June 30, 2023

Abstract

The SNOWNLP algorithm for sentiment assessment is used to evaluate Eastern Fortune stock bar posts to determine investor sentiment. A VAR-DCC-MGARCH model is initially constructed in order to study the relationship between investor sentiment and the financial time series of the SSE index price. A benchmark regression is built once more to study how investor sentiment impacts stock prices. According to a mediating route test, daily trading volume is a mediating element of investor sentiment on the price of the SSE index, which further finds that investor sentiment has a direct significant positive influence on the stock price. Finally, both the instrumental variable endogeneity test and the replacing variable robustness test reach the same conclusion and support the validity of the results.

Keywords

Investor Sentiment, Stock Price Index, SNOWNLP, Multivariate GARCH, Mediation Effect, Crawler

1. Introduction

China's securities market has evolved rapidly since its creation, and its market-operating mechanism has been consistently streamlined and enhanced, contributing to the rise of capital finance. On a yearly basis, the market has been subjected to strong shocks, with each anomalous movement being intimately tied to investor sentiment. This is owing to the market's inexperience, immaturity, and newness, as well as an ineffective market mechanism dominated by individual investors. Stock market shocks are exacerbated by the consequences of irrational investor sentiment fluctuations, which harm investors' interests and impede China's long-term, stable growth. As a result, it is critical to investigate how investor attitude impacts the prices of macrostock indexes in accordance with behavioral finance theory.

The diagram below displays the research methodology employed in this work. To begin, a Python crawler is used to collect all of the comment text data over the course of about 401 natural days. The number of favorable and negative comments is then effectively retrieved utilizing natural language processing techniques and text analysis. The smoothed time series of the investor sentiment index is created using both data and the investor sentiment construction formula. A DCC-MGARCH model is also developed to investigate the link between investor sentiment and the SSE Composite Index. Finally, an early influence of investor sentiment indicators on stock prices is investigated using a Granger causality test and impulse response analysis. Finally, acceptable control variables are chosen in order to create a regression model and prove the presence of a substantial connection between investor mood and the stock price of the SSE Composite Index. Along with the aforementioned main framework, this paper includes the steps of testing the mediating effect with total daily trading volume as the mediating variable, discussing endogeneity using the instrumental variable 2SLS method, and testing robustness using total daily trading volume as the replacement variable (Figure 1).

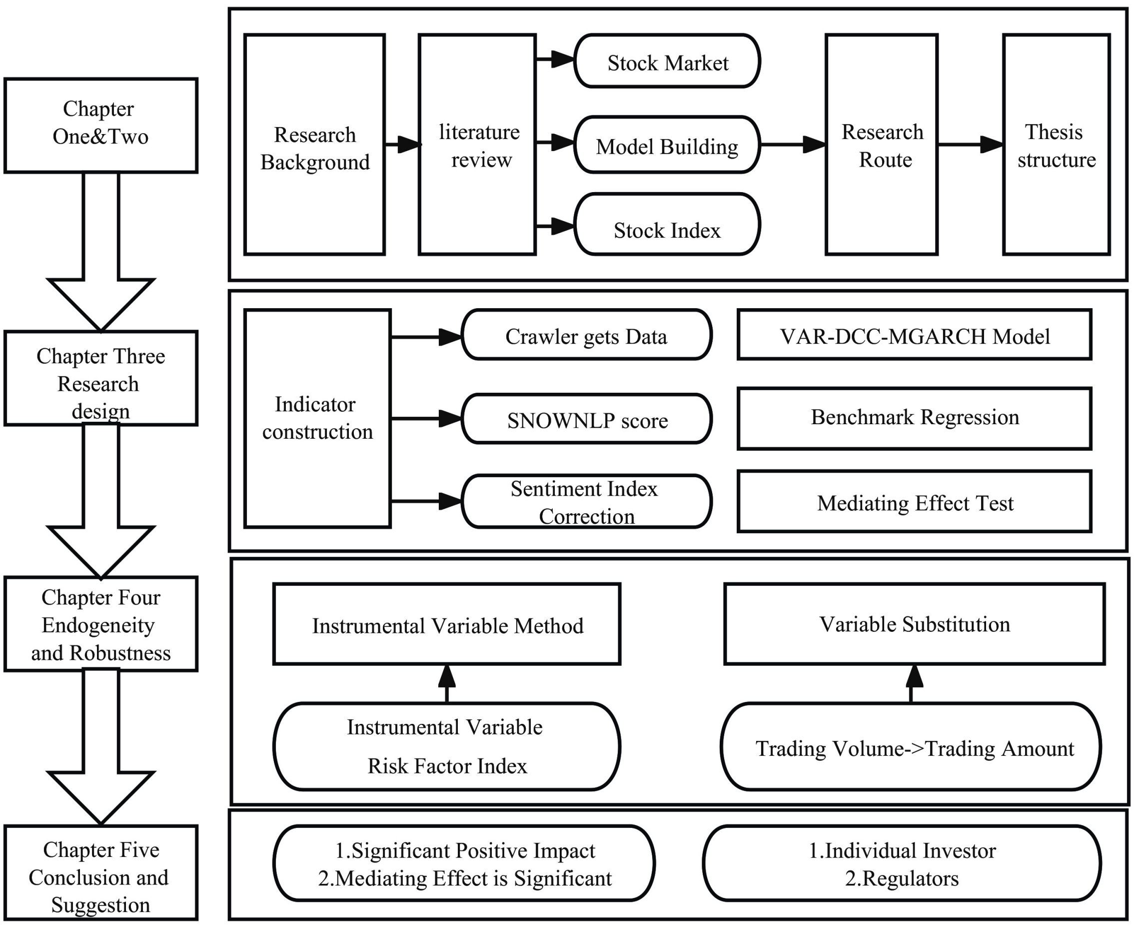

Figure 1. Research roadmap, from author.

Following the above procedure, the results gained in this study may be divided into two major points: 1) The investor sentiment index has a considerable positive effect on the share price of the SSE index. 2) The findings of the mediating effect study show that the trading volume effect, which acts as a mediating variable, has a considerable impact on the SSE index closing price. This research implies that trading volume works as a buffer between investor mood and the closing price of the SSE index.

Crawlers and SNOWNLP are two cutting-edge data collection approaches employed in this work. The adjustment approach for the investor sentiment index, which uses a non-trading day weighted average method, is innovative. As the control variable, the risk factor index is the instrumental variable, and the exchange rate is the instrumental variable. The theoretical significance of this work is that measuring the investor sentiment index remains a difficult topic that has not yet been completely handled. This article utilizes cutting-edge technologies, such as crawlers and natural language processing, to mine and extract investor sentiment from stock bar posting themes, therefore improving the text mining approach. Initially, cutting-edge technology enables the production of totally automatic indicators expressing market sentiment without the need to devote a significant amount of people in identifying the attitude of investors hidden in the text. Furthermore, the indicators produced in this research are logical and scientific, and a similar method may be utilized to anticipate changes in stock indexes using the investor sentiment index. Finally, the sentiment index gives a precise reference meaning to an investor, allowing them to use stock market sentiment to intervene in risk and gain profits.

The thesis is divided into four pieces. The introduction provides a thorough explanation of the research context for this study, its significance, and a summary of the domestic and international literature on three topics: the relationship between investor sentiment and the stock market, model development and the impact of investor sentiment indices on stock indices, the research framework of this study, and its potential innovations. The second section of the research design covers the construction of investor sentiment indicators, which is divided into three main sections: using crawlers to collect data, SNOWNLP to identify the sentiment of the data, and using earlier formulas to construct indicators, outlining the procedures and innovations of indicator construction. The vector autoregressive model is built to generate impulse response analysis plots to investigate the influence relationship between the two; the Granger causality test is run on the smooth financial time series to obtain two-way Granger causality; and the DCC-MGARCH model is built to conclude that the correlation between the two is significant. Finally, a mechanism test is performed using trade volume as the mediating variable, and a baseline regression model is constructed with control variables included. The final section discusses the endogeneity test and the robustness debate. The validity of the results is proven by using the instrumental variables technique for endogeneity testing and the substitution variables approach for robustness testing. The fourth section is the findings and suggestions, which offer advice to both individual investors and regulators based on the two conclusions reached.

2. Literature Review

Several studies on investor sentiment have been conducted. Crawlers acquire online forum comment data for text recognition and set up sentiment dictionaries or keywords to determine positive, negative, and neutral sentiment, which are the two most common methods for providing indicators for investment sentiment. The number of comments received in each category is artistically transformed into sentiment indicators for stock indexes or individual stock profiles to see whether the results are substantial enough to draw conclusions. The second stage is to extract current indications using principal component analysis to generate an investor sentiment index for future modeling exploration.

Guo et al. (2017) uses an optimal path approach to dynamically analyze investor sentiment and the stock market to conclude that stock prices are more accurately predicted only when investors are highly concerned about the stock market. Ma et al. (2018) study the link between market mood and excess returns, demonstrating that ETF price deviations are substantially impacted by market sentiment and that the impact effect is amplified in times of financial difficulty.

Hu et al. (2018) used the SVAR model to investigate the link between investor sentiment and stock market returns of various sizes and discovered that there is an interaction between investor sentiment and stock market returns. Zhao & Li (2019) concluded that investor sentiment affects stock returns and that there is a two-way Granger causality between the two. He et al. (2019) used six distinct investor sentiment indices and utilized principal component analysis to create a composite investor sentiment index that corresponded to the Chinese stock market.

The domestically connected DCC-MGARCH model investigates the dynamic correlation between the two variables in terms of model design. Yin & Wu (2019) mine stock market broad market commentary data with crawler software, use natural language processing techniques for text analysis to create daily indicators of investor sentiment, and use the DCC-MGARCH model to conclude that investor sentiment and market excess return and liquidity are related. They differ in time; the dynamic correlation coefficient is often positive; and high-frequency daily data can capture more precise information than monthly data. Mei, Zhang, & He (2019) used network reptile technology to collect user discussion information from the mobile internet and study the impact of user emotions from the mobile internet on stock earnings. The empirical results revealed that mobile internet users' emotions tend to show positive optimistic emotions; at the same time, the more optimistic the mobile Internet user's emotions are, the higher the next share earnings.

In terms of the impact of investor sentiment indices on stock indices, Yao &

Liu (2022) investigate the spillover effect of investor sentiment fluctuations and the volatility of the SSE 50, CSI 300, CSI 500, and GEM indices and conclude that all four indices show significant time-varying dynamic correlation coefficients following optimistic media reports. Wang & He (2022) investigated the predictive potential of investor sentiment indices on the SSE index at various time scales, validated them, and confirmed that investor sentiment indices may reflect investors' sentiment information on stock market trend prediction. This demonstrates that research have shown that investor sentiment has a positive influence on stock indices and has a predictive function, and the greater the investor sentiment index, the higher the SSE stock index.

To summarize, there has been minimal study on building indicators using current stock bar commentary text data, no discussion of how investor sentiment influences stock indices, and no set rules for developing indices that more properly represent investor emotion. In this work, we use crawlers and sentiment analysis to develop a regression model that examines the macro-level impact of investor sentiment on stock indexes.

3. Study Design

3.1. Data Sources

This study examines the text information of daily post subjects in the stock bar of the Eastern Fortune SSE Index from February 18, 2022 to March 24, 2023, a period of 401 days, and builds the investor sentiment indicator using the three processes outlined below: In the beginning, the crawler approach is utilized to access and collect data from share bar postings. Secondly the crawled text content is scored using SNOWNLP in Python for sentiment text recognition, and the number of positive and negative posts with a threshold of 0.5 is tallied. Finally, the investor sentiment indicator is built using Antweiler & Frank (2004)'s enhanced algorithm. The Eastern Wealth SSE Index stock bar is chosen as the data source in this paper for three reasons: first, online stock bars provide shareholders with a channel platform to effectively transfer and interact with information in a timely manner; second, investor sentiment can be influenced by each other through the exchange of comments; and third, stock bar exchanges are consistent with the characteristics of individual investors, who are the mainstay in China. Second, with around 100 million daily views and tens of millions of daily active users, the online stock bar is currently the most popular stock exchange network for Chinese investors and can accurately and swiftly represent changes in investor mood. Finally, the enormous number of entries on the Eastern Fortune website and the extended retention duration of the stock bar might give useful data for this research. As a result, the title message of this stock bar is chosen as the source for the crawler to acquire text data in this study.

The web crawler's primary function is to extract the source code of Internet pages. The application is used to connect to the website, download the page's source code, and extract the required textual content. This research used Python to directly crawl important data through the requests module, parsing the original comment data saved as a csv. Date, post title, and a total of 894153 crawl postings are specific crawl fields.

There is no shortage of invalid information in stock bar posts, so this paper uses the following screening methods to cleanse the data and eliminate invalid data:

1) Delete re-posted content from other websites, e.g., text containing the word "re-post".

2) Delete official news and announcement messages.

The total number of posts remaining after one round of data cleaning is 839,173. The table below shows the SSE Composite Index share bar price information (Table 1).

3.2. Construction of Investor Sentiment Indicators

Text sentiment analysis is a type of natural language processing (NLP) that involves assessing and summarizing subjective text with emotive overtones. This section is an important aspect of developing investor sentiment indicators since it provides a direct supply of data for the next stage of indicator development. This work employs the SNOWNLP Python package for text sentiment analysis, which includes a trained sentiment dictionary and a simple Bayesian model as its method. The model assumes that the feature words in the text are independent of one another and predicts the posterior probability that a piece of text's content fits into each category. Finally, because the return value reflects the chance that the text contains positive emotion, a return value closer to 1 suggests a higher propensity toward positive sentiment, while a return value closer to 0 indicates a stronger tendency toward negative sentiment.

In this study, the guidelines for assessing positive and negative sentiment are as follows: a return value more than 0.5 is a positive post, a return value equal to 0.5 is a neutral post, and a return value less than 0.5 is a negative post. The neutral text largely consists of terms like "retweet", "dive" and unidentified numbers, which SNOWNLP detects as 0.5. Positive communications generally use phrases like "soaring" and "bull market is on its way", whereas negative texts mostly use words like "high opening and low closing" and "Black Thursday". The table below shows the number of favorable and negative posts. The table below contains data on the number of good and negative posts (Table 2):

Table 1. SSE composite index share bar information, from author.

Information dimension	Related data
Time spacing	18 February 2022-24 March 2023
Total number of posts	839,173
Highest number of posts in a single day	18,707
Minimum number of posts in a single day	53

Table 2. Text recognition information, from author.

Emotions point to	Quantity information	Indicator Code
Positive Posts	411,354	Pos_cnt
Neutral posts	10,295	Neu_cnt
Negative Posts	417,524	Neg_cnt
Total number of posts	839,173	Total_cnt

After obtaining the number of daily positive and negative posts, the index is constructed according to the improved formula for constructing investor sentiment indicators by Antweiler & Frank (2004) with the following formula:

$$S_t = \ln\frac{1+\text{Pos\_cnt}}{1+\text{Neg\_cnt}} \approx \frac{\text{Pos\_cnt}-\text{Neg\_cnt}}{\text{Pos\_cnt}+\text{Neg\_cnt}} * \ln\left(1+\text{Total\_cnt}\right)$$

Because the closing price of the SSE Composite Index is a time series of trade days and investor sentiment is a time series of all days, the time dimension correction of the two models is critical for accurate assessment of investor sentiment. As a result, the correction approach employed in this work for the adjustment of the investor sentiment index for weekend closures, holiday closures, and temporary closures is as follows:

$$S_{\text{open}} = \frac{\sum_{i=1}^{k} S_{t-i} + S_t}{k+1}$$

That is, if the market closes over the weekend and reopens on Monday, the investor sentiment index for Monday is adjusted to the arithmetic mean of the investor sentiment indicators for Saturday, Sunday, and Monday in order to avoid losing information on the fluctuations of the investor sentiment index during the closed period. Finally, the table below demonstrates that the minimum value of the daily number of positive and negative postings is very near, both around 700, for the time series of investor sentiment that has been produced for simple descriptive statistical analysis. However, when comparing the maximum value and standard deviation of the two, the negative posts are significantly larger than the positive posts. In terms of mean values, the positive and negative post values are similar (**Table 3**).

3.3. Test of Smoothness

This part begins by charting the created time series normal distribution and time series diagrams of investor sentiment and SSE index prices, then differencing the shaky financial time series and performing Granger causality tests. Second, a vector autoregressive model is built, and impulse response analysis is used to investigate the effect of a one-unit shock on the two's dissimilar motions. Finally, the link between the investor sentiment index and the SSE index is investigated using a simple least squares linear regression. This section aims to investigate the statistical significance of the causal relationship between the two through causality

Table 3. Descriptive statistics on investor sentiment, from author.

	Maximum value	Minimum value	Average	Standard deviation
Positive Posts	8557	725	1534.903	863.877
Negative Posts	9862	722	1557.925	1008.950

testing, the relationship between the two time series through vector autoregressive modeling, the impact of exogenous shocks on the trend of the two through impulse response analysis, and the linearly significant relationship between investor sentiment and the price of the SSE Index through a final simple linear regression.

1) causality test analysis

This study uses a normal distribution chart to depict the distribution of investor mood indicators as well as the time series of the SSE closing price. The investor sentiment index has a clear normal distribution, with the majority of values concentrated around −0.1 - 0.1 and relatively few values less than −0.3 and larger than 0.4 (**Figure 2**).

Furthermore, plotting the time series of the SSE closing price and investor sentiment clearly shows that the investor sentiment index's trend is likely to overlap with the SSE's high and low points, and that the investor sentiment index fluctuates in the range of −0.3 - 0.45 or so, with the SSE fluctuating in the range of 2900 - 3500 points or so. It can be observed that the investor sentiment index has a definite propensity to shift throughout time cycles with a high degree of volatility. The goal of the simple moving average operation used here with exponential smoothing is to remove the influence of temporal trends on the time series of investor sentiment, as illustrated in the graph below on the right. This graph shows that the volatility patterns between the two are continually converging, with the highs and lows largely overlapping (**Figure 3**).

To prevent regression difficulties, the starting variables of investor sentiment, investor sentiment indicator, and SSE Composite Index stock price, whose samples are time series, must be evaluated for smoothness. As a result, an ADF test is run to convert the unstable time series into a steady time series. The original series of the SSE index is unsteady, but its first-order differential series is steady, so the Granger causality test uses its first-order differential series as the original data for investor sentiment and the closing price of the SSE index, as well as its differential post-test p-value (**Table 4**).

Following the Granger causality test, it is established that the investor sentiment index and the first-order difference of the SSE index series are Granger causes of each other, indicating that they exhibit mutual causation in a statistical sense. Other factors may still be influencing the shift between the two, but this research focuses solely on the link between the two. The following are the particular hypothesis and test p-values, which all reject the original hypothesis (**Table 5**).

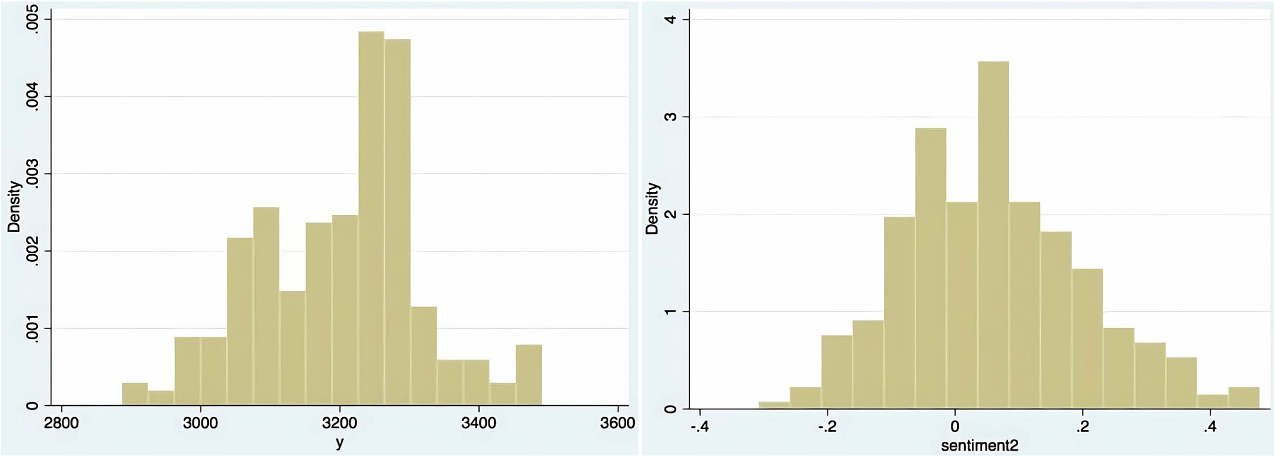

Figure 2. Normal distribution of investor sentiment and SSE index, from author.

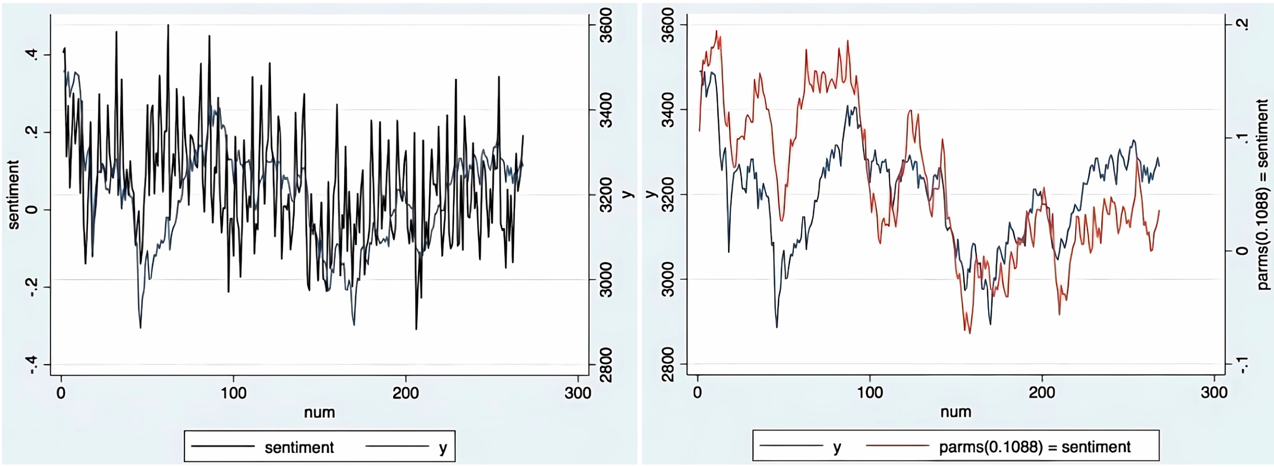

Figure 3. Time series diagram, from author.

Table 4. Time series difference, from author.

Variable name	Variable Codes	*p*-value
Investor sentiment	sentiment	0.000***
Investor sentiment first order differential	Diff_1_sentiment	0.000***
Investor sentiment second order differential	Diff_2_sentiment	0.000***
SSE	*y*	0.187
SSE First Order Difference	Diff_1_y	0.000***
SSE Second Order Difference	Diff_2_y	0.000***

Table 5. Results of Granger's causality test, from author.

Original assumption H0	*p*-value
Granger reason why sentiment is not Diff_1_y	0.090*
Granger's reason why Diff_1_y is not a sentiment	0.001***

2) VAR model and impulse response analysis

Since investors and stock indices are endogenous and lagging, this paper uses VAR to test the relationship between them in order to investigate their mutual influence.

Before building the vector autoregressive model, it is necessary to confirm the optimal lag term order using stata to calculate the values of different orders of LR, FPE, AIC, HQIC, and SBIC and confirm the optimal lag term order using the principle of the most significant indicator of the smallest order. The table is as follows (Table 6).

After constructing the VAR model, the *p*-value of the investor sentiment indicator relative to the first-order difference SSE index is 0.09, which is significant at the 10% confidence level. Secondly, the unit roots are plotted as follows to show that they are all located inside the unit circle, allowing for the construction of the impulse response function in the next step (Figure 4).

The impulse response function examines the size of a unit shock's influence

Table 6. Lagging term indicators, from author.

lag	LR	FPE	AIC	HQIC	SBIC
0	--	21.559	8.747	8.757	8.774
1	43.038	18.870*	8.613*	8.646*	8.695*
2	3.8123	19.173	8.629	8.684	8.765
3	4.2913	19.445	8.643	8.720	8.833
4	11.257*	19.206	8.631	8.729	8.875

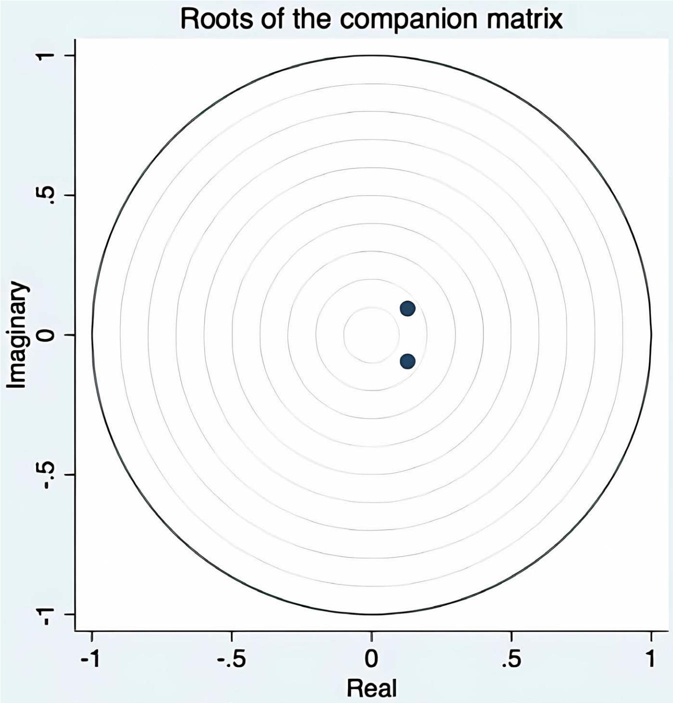

Figure 4. Unit root test, from author.

on the trend of each variable. Stata is used to plot the impulse response function, which is saved in myfile, and the impulse response model for 10 periods is chosen for this part. The findings are converging in all four graphs, with encouraging results for additional research. The top left and bottom right graphs of investor sentiment and first order differential SSE index price on own shocks show that investor sentiment is more path dependent than SSE index price because investor sentiment trends are less variable than stock index prices, which are more stable. The top-right panel depicts the SSE price's reaction to the investor mood index. It can be observed that a one-unit change in the investor sentiment index creates a negative fluctuation in the SSE index, ranging from 0 to -3, following which the SSE price climbs back to 0 after two periods and returns to a stable state. The graph on the left depicts the investor sentiment index's reaction to the SSE price, with a unit of price fluctuation causing the investor sentiment index to move moderately downwards in period 1 and drastically downwards in periods 2 - 4, from around 0.04 to 0 (**Figure 5**).

3) DCC-MGARCH model analysis

The multivariate GARCH model is intended to study the dynamic correlation between a number of time series. The DCC-MGARCH model is used in this article to explore the relationship between the investor sentiment index and the SSE share price. First, the ARCH effect test is performed, and the DCC-MGARCH model is built if the findings are significant. Again, the model is built to observe the magnitude of each indicator's coefficients and to do dynamic correlation analysis.

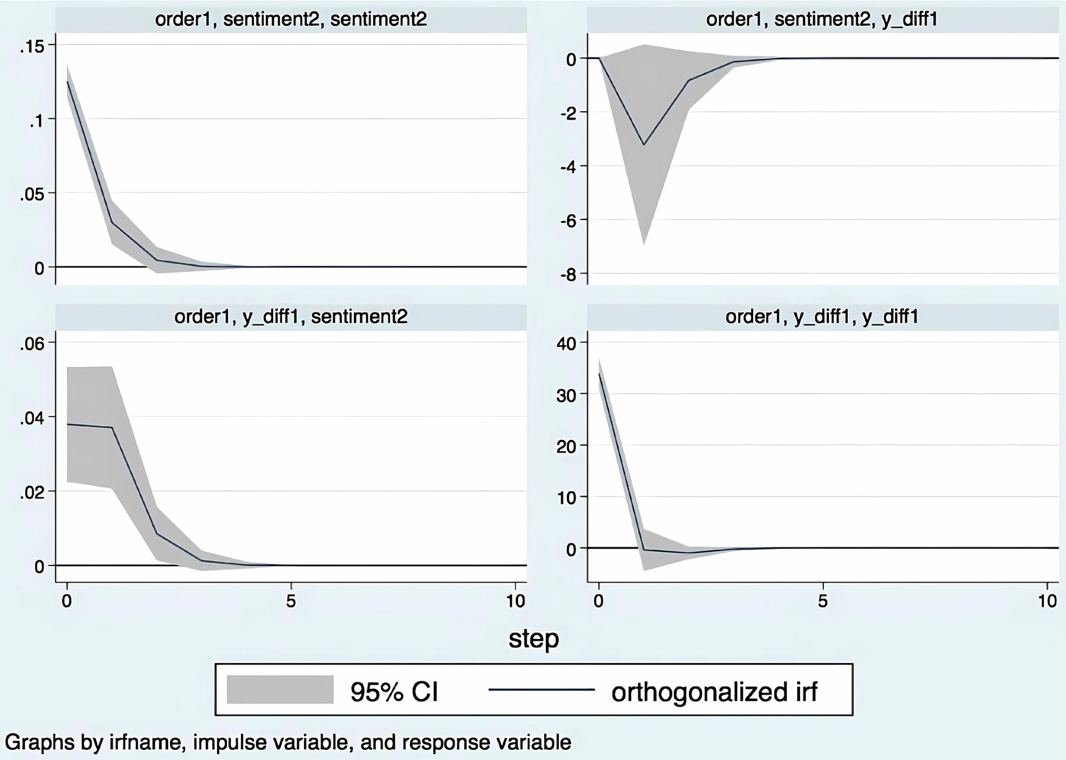

Figure 5. Impulse response diagram, from author.

A 6th-order ARCH effect test is conducted, and the results of the indicator values shown below are obtained. It can be seen that all 6th-order effects are significant at a 5% confidence level, with the first 4th-order tests all being significant at a 1% confidence level (Table 7).

A DCC-MGARCH model is constructed to determine the values of the correlation indicators between the independent variable investor sentiment index and the dependent variable SSE Composite Index stock price as shown in the table below, where, and a small value indicates that the last index shock has a small impact on the correlation coefficient, and a larger value indicates a strong persistence feature between the two time series (Table 8).

3.4. Baseline Regression

This paper establishes a regression model for the main regression test by selecting the overnight shibor rate, the daily exchange rate, the lagged first-order term of investor sentiment, and the lagged second-order term of investor sentiment as control variables and constructing the regression model shown below:

$$y = \alpha_0 + \beta_1 Se + \sum \beta_k C + \gamma$$

where y represents the SSE share price; represents the intercept term; β are the coefficients; C is the series control variable; and is the disturbance term.

The following table shows the descriptions and definitions of the main variables (Table 9):

For each variable, descriptive statistical analysis and correlation tests are performed, and the descriptive statistical graphs are provided below (Table 10). The plots demonstrate that 1) the mean value of the dependent variable, SSE Index

Table 7. ARCH effect results, from author.

lags(p)	chi2	Prob > chi2
1	10.770	0.002
2	13.344	0.001
3	14.081	0.003
4	14.369	0.006
5	14.277	0.014
6	15.715	0.015

Table 8. Results of the DCC-MGARCH model, from author.

Dependent variable	Independent variable	Coefficient values	p-value
y	L2.sentiment	0.89860	0.954
sentiment	L2.y	0.00017	0.023
Corr(y, sentiment)		0.55675	0.000
Lambda1		0.00495	0.596
Lambda2		0.95658	0.000

Table 9. Definition table for variable classification, from author.

Variable classification	Name	Symbols	Definition of variables
Explained variables	SSE Share Price	y	Changes in the SSE generally represent broad market point movements
Explanatory variables	Investor Sentiment Index	Se	Constructed using the formula
Control variables	Overnight shibor	SH	Shanghai Interbank Offered Rate
	Exchange rates	IR	Renminbi to US Dollar Daily Exchange Rate
	First-order lag in investor sentiment		First-order lags of the explanatory variable Se
	Second order lag in investor sentiment		Second-order lags of the explanatory variable Se

share price, is 3200.570, indicating a right-skewed roughly normal distribution. 2) The ADF test demonstrates that all time series are smooth except the financial time series of the SSE index, which is not. 3) According to the correlation coefficient table (Table 11), the dependent variable has a substantial positive connection with the independent variable, whereas the exchange rate level has a negative association with all variables.

As shown in the first column of Table 12 (values in brackets represent t-value and all charts are the same), the investor sentiment index has a substantial positive influence on the SSE stock price at the 1% confidence level even when the regression coefficients of the control variables are excluded. Through two regressions, the first two and third columns of the first and second order investor sentiment index, shibor, and exchange rate are gradually included as control variables. As the control variables are gradually introduced, the discernible coefficients grow and the degree of fit increases. This shows that investor mood has a substantial impact on the SSE (Table 12).

3.5. Intermediary Pathway

According to theory, investor sentiment has a direct link with the SSE Composite Index share price, with positive sentiment investors choosing to acquire shares, causing the share price to rise, and vice versa. As a result, as positive sentiment grows, the number of people who feel the stock price will climb and are considering buying the stock grows, as does the volume of trading and the quantity moved, leading the stock price to rise. In contrast, when negative sentiment grows, more individuals feel that the stock price will decrease and opt to sell the shares, resulting in an increase in trading volume and, eventually, a difference in the amount exchanged. As a result, the total daily trade volume may be used as a mediating variable to test for mediating effects, as shown below: The symbol for the mediating variable is VB, and the following mediating effect test model is developed:

$$D = \alpha_0 + \beta_1 Se + \sum \beta_k C + \gamma$$

$$M = \alpha_0 + \beta_1 Se + \sum \beta_k C + \gamma$$

$$D = \alpha_0 + \beta_1 Se + \beta_2 M + \sum \beta_k C + \gamma$$

Table 10. Descriptive statistics, from author.

Code	Average	Standard deviation	Minimum value	Maximum value	Skewness	Kurtosis	ADF	number of observations
y	3200.570	118.829	2886.426	3490.757	0.609	0.982	0.187	268
Se	0.058	0.143	−0.308	0.478	0.009	0.898	0.000	268
SH	1.461	0.392	0.441	2.459	0.046	0.170	0.000	268
IR	6.815	0.518	6.301	14.251	0.000	0.000	0.000	268
	0.058	0.143	−0.308	0.478	0.008	0.866	0.000	268
	0.057	0.143	−0.308	0.478	0.007	0.879	0.000	268

Table 11. Spearman correlation coefficient, from author.

Code	*y*	*Se*	SH	IR		
y	1.000					
Se	0.323**	1.000				
SH	0.252**	0.049	1.000			
IR	−0.492**	−0.357**	−0.319**	1.000		
	0.286**	0.340**	0.036	−0.368**	1.000	
	0.293**	0.155**	0.068	−0.365**	0.346**	1.000

Table 12. Regression results, from author.

Variable Codes		*y*	
Se	303.183*** (6.380)	206.966*** (4.350)	197.263*** (4.340)
		145.856*** (2.980)	126.489*** (2.690)
		194.457*** (4.150)	159.821*** (3.520)
SH			61.457*** (3.840)
IR			−34.818*** (35.910)
	0.133	0.218	0.294
F	40.710	24.410	21.650

where *D* represents the dependent variable SSE stock price; *M* represents the mediating variable VB; and *Se* represents the investor sentiment index.

First of all, Table 12 demonstrates that investor sentiment has a considerable positive influence on the SSE index, with a coefficient of 303.1831 at the 1% confidence level. The coefficients and regression results from Stata are displayed in the table below. The regression findings with the investor sentiment index as the independent variable and the SSE index price as the dependent variable have substantial non-zero coefficients, and the entire equation is considerably non-zero.

The logit regression and total effect regression findings are provided below, using the mediating variable as the dependent variable, investor sentiment as the independent variable, and a number of control factors. The significant values of a, b, and c for the total and indirect effects are verified by a Sobel test result of 2.999 over 1.96, indicating that the mediating variable is a completely mediated effect (Table 13).

4. Endogeneity Discussion and Robustness Test

4.1. Instrumental Variables Test

Because this article investigates the link between the investor sentiment index and the closing price of the SSE, the instrument variable used should be significantly associated with the investor sentiment index but not causally related to the SSE. As a result, in this work, the liquid market capitalization-weighted risk factor index (RF) for the financial industry is used as the instrumental variable. As a measure of sector risk, the risk factor index indicates the amount of risk and the quantity of investment possibilities, whereas investor sentiment measures investor mood toward the market. When the risk index rises, investors get concerned and have negative attitude, causing the investor sentiment index to decline. It, on the other hand, increases. The Risk Factor Index is an exogenous variable that influences investor sentiment and hence indirectly impacts SSE stock price fluctuations, and its usage as an instrumental variable successfully controls for endogeneity difficulties.

This paper uses a two-stage 2SLS regression to address possible endogeneity, and the following table reports the results of the two-stage regression and the instrumental variable test. The first column shows that the instrumental variable has an F-value of 8.05 < 10, which is a weak instrumental variable. A 2SLS regression test is conducted on this, and the first-stage results show that the relationship between the instrumental and endogenous variables is significant at the 1% confidence level. The second stage results show that the predicted value of investor sentiment in the first stage and all exogenous variables are significant at

Table 13. Results of the test for intermediate effects, from author.

Variable Codes	ln_y	ln_VB	ln_y
Se	0.095*** (0.015)	0.463*** (0.083)	0.077*** (0.015)
VB			0.038*** (0.011)
Control variables	Control	Control	Control
	0.131	0.105	0.171
F	40.220	31.110	27.380

Sobel test value: 2.999 > 1.96.

the 1% level, indicating that the relationship between the endogenous and exogenous variables is significant, which in turn indicates that the results obtained by the 2SLS method are reliable (**Table 14**).

4.2. Variable Substitution

Since total daily trading volume and total daily trading amount have similar trends in terms of the impact of investor sentiment on the closing price of the SSE index, the robustness test replaces total daily trading volume with total daily trading amount TA for the mediation effect test, and the test results are shown in the table. Using the same methodology to calculate the total and indirect effects

Table 14. 2SLS regression results, from author.

Variable Codes	(1) *Se*	(2) *y*
RF	1.184*** (3.670)	
Se		197.263*** (4.340)
	0.222*** (3.600)	126.489*** (2.690)
	0.010 (0.170)	159.821*** (3.520)
SH	−0.022 (−0.960)	61.457*** (3.840)
IR	0.021 (1.190)	−34.818*** (−2.820)
	0.134	0.243
F	8.050	20.930

Table 15. Results of tests for mediating effects of variable substitution, from author.

Variable Codes	ln_y	ln_TA	ln_y
Se	0.062*** (0.014)	0.324*** (0.074)	0.046*** (0.014)
ln_TA			0.048*** (0.012)
	0.040*** (0.015)	0.233*** (0.076)	0.028** (0.015)
	0.050*** (0.014)	0.177** (0.073)	0.042*** (0.014)
Control variables	Control	Control	Control
	0.291	0.250	0.334
F	21.320	17.320	21.62

Sobel test value: 3.003 > 1.960.

and derive the sobel test values and *p*-values, even though the a and c values are significant, the b value and the *p*-value of the final sobel test are significant, confirming the robustness of this study's findings (**Table 15**).

5. Conclusion and Recommendations

This study creates an investor sentiment index using crawlers to gather post-title messages from the Eastern Fortune SSE Index bar and text sentiment analysis using the SNOWNLP scoring model. The relationship between the SSE stock price and the investor sentiment index, as well as the dynamic correlation between them, are examined using a VAR-DCC-MGARCH series model. Once more, appropriate control variables are chosen, and regression models are created to investigate the mechanism of influence between the two.

The empirical findings demonstrate that 1) the investor sentiment index significantly affects the share price of the SSE Index, i.e., the higher the investor sentiment, the higher the SSE Index share price. 2) The results of the mediation effect test reveal that the total daily trading volume, as a mediator, significantly mediates the investor sentiment index's significant effect on the SSE closing price. The total daily trading volume is substituted for the variable in the robustness testing in this paper, which also uses instrumental variable tests to investigate the endogeneity problem, and it is discovered that the aforementioned conclusions are still true.

The conclusions of this paper realistically demonstrate that the investor sentiment index extracted from the Eastern Fortune stock bar forum can effectively predict fluctuations in stock index prices, providing insights for both individual investors and regulators. This paper theoretically enriches the number of papers related to the influence of investor sentiment on the SSE index in the Chinese market. This paper provides appropriate advice to individual investors and system regulators by demonstrating the mechanism of significant positive influence on the index of individual investor sentiment, and reflects the economic fundamentals from the side if there is major negative news, policy austerity, or increased geopolitical risk, then through the emotional transmission of individual investors, resulting in the existence of the index reliability decline. Furthermore, this paper investigates not only the attitude of ordinary investors, but also the sentiment of institutional investors, which has a significant impact on the growth in the index stock price. Finally, the data in this article only picked up the forum post title for a year, with data calibration discarded just 268 days after the non-trade date, and the existence of this article is also limited by a lack of sample. 1) Based on the investor sentiment pointers displayed in the forum, individual investors can assess the general macro index movements. If the investor sentiment index is at a high level, it suggests that the market is experiencing overly optimistic psychology, which should be suppressed to maintain a calm and logical psychological analysis of market conditions. Quality stocks with long-term investment value can be considered instead. 2) The regulator can observe the trend of the investor sentiment index to understand the sentiment of market participants so that timely regulatory measures can be taken.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Antweiler, W., & Frank, M. Z. (2004). Is All that Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance, 59,* 1259-1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Guo, K., Sun, Y., & Qian, X. (2017). Can Investor Sentiment Be Used to Predict the Stock Price? Dynamic Analysis Based on China Stock Market. *Physica A: Statistical Mechanics and Its Applications, 469,* 390-396. https://doi.org/10.1016/j.physa.2016.11.114

Ma, J. A., Ho, K. C., Yang, L., & Chu, C. C. (2018). Market Sentiment and Investor Overreaction: Envidence from New York Listed Asian Country Exchange Traded Funds. *Emerging Markets Finance and Trade, 54,* 2455-2471. https://doi.org/10.1080/1540496X.2018.1464907

He, G., Zhu, S. Z., Gu, H. F., & Wu, X. F. (2019). A PCA Method to Build Chinese Stock Market Compound Sentiment Index. *Journal of Donghua University (Natural Science), 45,* 319-326.

Hu, Z. Y., Ye, Y. Y., & Zhang, J. X. (2018). The Interactive Influence between Investor Sentiment and Stock Market Returns of Different Scale: Based on Empirical Study of SVAR Model. *Journal of Guangzhou University (Natural Science Edition), 17,* 8-20.

Wang, N., He, Y. Y., & Liu, L. (2022). Research on the Construction and Effectiveness of Investor Sentiment Index—Based on Emotional Analysis Oriental Wealth Stock Bar Posts. *Price Theory and Practice,* 146-151. https://doi.org/10.19851/j.cnki.cn11-1010/f.2022.11.353

Yao, H. X., & Liu, M. H. (2022). Research on Media Reports, Investor Sentiment and Stock Index Volatility. *Price: Theory and Practice,* 150-153+207. https://doi.org/10.19851/j.cnki.cn11-1010/f.2022.09.427

Yin, H. Y., & Wu, X. Y. (2019). Daily Investor Sentiment, Excess Return and Market Liquidity—Research on Time-Varying Correlation Based on DCC-GARCH Model. *Journal of Beijing Institute of Technology (Social Sciences Edition), 21,* 76-87+114. https://doi.org/10.15918/j.jbitss1009-3370.2019.1776

Mei, L. X., Zhang, C., & He, L. (2019). Investor Sentiment and Stock Returns: An Empirical Study from Mobile Internet. *South China Journal of Economics,* 36-53. https://doi.org/10.19592/j.cnki.scje.360011

Zhao, S. Y., & Li, J. (2019). Effect of Investor Emotions on Rate of Return on Stock. *The Economist,* 74-75+78.