Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 13
BMMFT 2024

The Impact of Investor Sentiment on Stock Returns

Xin Chen *

School of Economics and Management, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, P. R. China

* Corresponding Author Email: xin_chen1021@163.com

Abstract. This paper aims to explore the relationship between investor sentiment and stock returns. An investor sentiment index was constructed using principal component analysis, and its reliability was verified. The empirical research results indicate that the returns of the CSI 300 Index are significantly positively correlated with investor sentiment, meaning that when the sentiment index increases, returns also rise; conversely, when the sentiment index decreases, returns fall. Similarly, a significant positive correlation exists between investor sentiment and the returns of the SSE 50 Index, CSI 500 Index, Shanghai Composite Index, and Shenzhen Component Index. Further construction of a VAR model and impulse response analysis revealed the mutual influence between SENT (sentiment index) and stock returns. The study found that SENT has a more substantial impact on the indices in the short term, but this effect gradually weakens and converges to zero over time. This suggests that investors are more influenced by sentiment in the short term, but as market information increases, the impact of sentiment on the market diminishes. This research is of great significance for understanding the influence of investor sentiment on the stock market.

Keywords: Investor Sentiment, Principal Component Analysis, CSI 300 Index, Single-Factor Regression, VAR Model.

1. Introduction

Behavioral finance argues that the traditional financial theories, such as the Efficient Market Hypothesis and the Rational Investor Hypothesis, are insufficient to fully explain the behavioral phenomena observed in financial markets. In reality, investors' behavior in financial markets is often not entirely rational. Taking China's stock market as an example, we can observe that investors are frequently influenced by emotions and cognitive biases, leading them to engage in behaviors such as following rumors, blindly chasing trends, and seeking speculative opportunities. These behaviors do not conform to the "rational person" model assumed in traditional financial theories. Therefore, the development of behavioral finance has gradually integrated perspectives from psychology and finance to explain these phenomena in financial markets.

Behavioral finance suggests that investors' decisions are often influenced by their past investment experiences and the various information they acquire, and these cognitive biases affect their judgment of the market. Moreover, these biases can lead to an emotional contagion effect, where optimistic or pessimistic emotions are amplified, thereby increasing the volatility of the stock market.

This study is based on the investor sentiment theory within behavioral finance, aiming to explore the impact of investor sentiment on financial markets. According to this theory, investor sentiment is considered one of the key factors causing anomalies in financial markets. When making investment decisions, investors are not only processing the information they collect but are also going through a process of psychological cognition and emotional fluctuation. Investor sentiment reflects their views on market trends, which indicate their behavioral tendencies and investment willingness, and may translate into actual investment behavior. Given the contagious nature of investor sentiment, when investment inclinations excessively convert into actual investment actions, it can impact market trends and cause market returns to deviate from normal and rational levels.

2. Analysis of the Impact of Investor Sentiment on Stock Returns

Investor sentiment refers to the deviation between rational and irrational expectations when investors forecast the future prices of stocks. According to previous research conducted by scholars both domestically and abroad, it is reasonable for different countries and research groups to select different variables as proxies for investor sentiment. The most direct definition is investors' overall perception of the stock market, whether optimistic or pessimistic.

Typically, the proxies used to measure investor sentiment can be categorized into three types. The first approach focuses on individual investors by analyzing the frequency of keywords on search engines such as Google, Baidu, and social media platforms like Weibo, or by extracting information from posts and blog articles on sites like Eastmoney. This type of data reflects the emotions and views of individual investors regarding market trends. The second approach considers institutional investors by analyzing trading data to reflect their sentiment. Common indicators include turnover rates, the number of new accounts opened, and trading volume, which can demonstrate the level of market participation and trading activity of institutional investors, thereby indirectly indicating their sentiment and expectations. The third approach involves factors that have less direct correlation with the economy and finance, such as weather changes or politically sensitive topics. These factors can affect investors' emotions and decisions, thereby influencing the market.

In behavioral finance research, earlier scholars believed that there is a significant correlation between the sentiment of each investor in the stock market and stock returns. As a result, the momentum factor has been introduced into behavioral finance capital asset pricing models to measure the impact of market sentiment. This indicates that fluctuations in investor sentiment may lead to irrational market swings and deviations from normal price levels.

In their 2006 study, Baker and Wurgler comprehensively considered both subjective and objective factors to construct new indicators for studying investor sentiment. They found that investor sentiment exerts varying degrees and directions of impact on stock returns across different time periods. In the Chinese academic community, scholars have conducted extensive empirical research on the influence of market anomalies in the Chinese stock market on investors. The studies by Chen Liuqin and Liu Renhe (2005), and Wu Yanran and Han Liyan (2007), showed that investor sentiment significantly affects both current and future expected returns in China's A-share market. Additionally, Zhang Qiang et al. (2007) categorized participants in the A-share market into institutional and retail investors, studying their investor sentiment separately and calculating sentiment indices for each type of investor. The results revealed that institutional investors have a more substantial impact on the stock market than smaller investors.

These empirical studies suggest that investor sentiment plays an important role in the Chinese stock market, influencing both market volatility and returns. Different types of investor sentiment may have varying effects, making it necessary to consider sentiment factors when researching and predicting market behavior.

3. Method and Process of Constructing the Investor Sentiment Index

3.1. Data Sources and Variable Selection

3.1.1. Sample Selection

This study uses the CSI 300 Index as the research subject, with the comprehensive investor sentiment index (including the Consumer Confidence Index, number of new account openings, monthly market trading volume, turnover rate, and price-earnings ratio) serving as the predictor variables. The sample data consists of monthly observations, sourced from the RESSET database, covering the period from January 2006 to March 2023, providing 207 data points for each variable.

3.1.2. Variable Selection

(1) CSI 300 Index

The CSI 300 Index is a financial indicator jointly compiled and published by the Shanghai and Shenzhen Stock Exchanges on April 8, 2005, to reflect the performance of 300 large-cap stocks listed on both exchanges. This index is considered a key indicator of the overall trend of the Shanghai and Shenzhen stock markets, as its constituent stocks represent a wide range of high-liquidity and representative stocks in the market. The CSI 300 Index not only reflects the investment returns of market institutions but also accurately indicates the overall performance of the Shanghai and Shenzhen stock markets. Historically, the index has shown a high correlation with the Shanghai Composite Index, with its constituent stocks mainly composed of leading stocks from various industries, displaying a stable upward trend.

Therefore, the CSI 300 Index is widely used to evaluate investment performance and serves as a foundation for index-based investment and derivative innovation in fund management. Additionally, it is used as the underlying asset for exchange-traded funds (ETFs) and index futures contracts, meeting investors' needs for diversified investment and risk reduction. The sectoral distribution of the CSI 300 Index closely matches the overall industry distribution in the market. Its close correlation with the international markets and macroeconomic factors allows it to accurately reflect changes in the Chinese stock market. As a result, the CSI 300 Index holds significant importance in investment decision-making and market analysis.

The returns of the CSI 300 Index are calculated using closing prices: $R_t = 100 * ln(P_t/P_{t-1})$. As shown in Table 1, the results of the Augmented Dickey-Fuller (ADF) test indicate that the returns series of the CSI 300 Index is stationary, making it suitable for targeted prediction and analysis.

Table 1. Testing the Stability of CSI 300 Index Returns

	t-Statistic	ADF（Prob.*）
Returns	-12.74721	0.0000

(2) Consumer Confidence Index (CCI)

The Consumer Confidence Index reflects consumers' views on the current social situation, as well as their subjective perceptions and satisfaction levels regarding economic prospects, quality of life, income levels, employment conditions, consumption psychology, and income expectations. It directly indicates the degree of satisfaction that consumers and the public feel towards the socio-economic situation, making it a crucial measure of investor sentiment. In China’s securities market, where individual investors dominate, the CCI serves as a reliable proxy for capturing investor sentiment. Additionally, because investor confidence index data is often difficult to obtain, this study uses the CCI as one of the factors influencing investor sentiment. By observing changes in the CCI, we can gain insights into the expectations of both consumers and investors regarding the economy and market, thus predicting market trends and investor behavior. In this study, the logarithm of the CCI is used.

(3) Turnover Rate (TURN)

The turnover rate refers to the ratio of the number of shares traded within a given time period to the total number of shares in circulation. It reflects the degree of trading activity among investors. When investor sentiment is optimistic, they are more likely to actively engage in stock trading, resulting in a higher turnover rate. Conversely, when investor sentiment is pessimistic, they may prefer to hold onto stocks or adopt a wait-and-see approach, leading to a lower turnover rate. Thus, the turnover rate is often regarded as an important indicator of investor sentiment. By monitoring changes in the turnover rate, we can reveal investors' trading intentions and sentiment towards the stock market, providing a basis for market trend analysis and predictions.

(4) Market Trading Volume (MMV)

Trading volume is an indicator that reflects the level of investor participation. When investor sentiment is high, trading volume tends to increase significantly. In the current context, observing the pace of IPOs (initial public offerings) and refinancing activities reveals that China's stock market is rapidly expanding, with market size continuously growing. To eliminate the impact of the expansion of the A-share market on trading volume, this study uses the ratio of the monthly trading amount of the CSI 300 Index to the monthly circulating market value. By dividing the monthly trading amount by the market value, we can remove the influence of market scale and more accurately reflect the level of investor activity and sentiment. This adjustment helps the trading volume better reflect changes in investor sentiment, aiding researchers in understanding the relationship between market participation and investor sentiment.

(5) Number of New Investor Accounts (NNA)

According to the research by Du Mian and Zhao Yurong (2016), the number of new investor accounts (NNA) is positively correlated with the stock price index, with a significant correlation. This means that when the number of new accounts increases, the stock price index usually rises, and vice versa. Therefore, using NNA as a proxy indicator to measure investor sentiment is reasonable in this study. An increase in the number of new accounts indicates heightened attention and participation in the stock market, reflecting a positive investor sentiment. Conversely, a decline in new account numbers suggests reduced investor interest in the stock market, potentially indicating a more pessimistic investor sentiment. Thus, using NNA as a proxy indicator helps analyze and study investor sentiment.

(6) Price-Earnings Ratio (PE)

The price-earnings ratio is the ratio of a stock’s market price to its earnings per share. It is a classic measure used to assess stock performance and is widely applied in gauging investor sentiment. A higher PE ratio often indicates that stocks are overvalued and that the market is overheated, while a lower PE ratio suggests that stocks are undervalued or that the market is cooler. According to research by Zhang Zongxin and Wang Hailiang, the PE ratio was chosen as one of the proxy variables for constructing the investor sentiment index, and the effectiveness of this method was confirmed. This implies that the PE ratio can, to some extent, reflect changes in investor sentiment. By observing changes in the PE ratio, we can infer investor evaluations of stocks and their sentiment status, thus enabling analysis and prediction of market trends. Therefore, including the PE ratio as one of the proxy variables for investor sentiment in the study is theoretically and empirically supported.

3.2. Descriptive Statistical Analysis

The descriptive statistical analysis for the selected variables is presented in the table below:

Table 2. Descriptive Statistical Analysis of Variables

	Mean	Median	Standard Deviation	Minimum	Maximum
Returns	1.0494	0.9106	9.6030	-29.9088	69.3532
CCI	4.6889	4.6826	0.0880	4.4485	4.8489
TURN	16.1929	14.0107	8.5358	6.0386	56.0428
MMV	22.8734	22.9270	0.6088	21.2958	24.5410
NNA	13.8849	13.9136	0.7225	11.6596	16.0043
PE	29.8471	27.1028	9.1725	18.1271	73.4102

3.3. Construction of the Composite Investor Sentiment Index

First, a composite investor sentiment index is constructed using principal component analysis (PCA) to aggregate the various investor sentiment indicators.

Table 3. Investor Sentiment Indicator Source

Consumer Confidence Index	CCI	Logarithm of the Monthly Consumer Confidence Index Published by the Statistics Bureau
Market Turnover Rate	TURN	(Shanghai and Shenzhen 300: Monthly Market Turnover / Average Total Market Value of the Last Two Months) * (Average Trading Days per Month / Total Trading Days in Each Month)
Market Trading Volume	MMV	Ratio of Monthly Trading Volume to Average Monthly Market Capitalization of the Shanghai and Shenzhen 300
Number of New Investor Accounts	NNA	Logarithm of the Number of New Accounts Opened Each Month
Price-to-Earnings Ratio (P/E Ratio)	PE	Ratio of Stock Market Price to Earnings Per Share (P/E Ratio)

Based on the methods of Baker and Wurgler (2012) and He Shasha (2018), this study selects five proxy variables to measure investor sentiment, including current and lagged values. The results indicate that these variables all have an impact on investor sentiment. To construct a comprehensive investor sentiment index, this study conducts principal component analysis (PCA) on these ten variables.

Table 4. Results of the Appropriateness Test for Principal Component Analysis of SENT

KMO Measure of Sampling Adequacy		0.654
Bartlett's Test of Sphericity	Approximate Chi-Square	3544.162
	Degrees of Freedom	55
	Significance	0.0040

First, SPSS software is used to conduct the KMO (Kaiser-Meyer-Olkin) and Bartlett tests on the selected proxy variables. These tests help analyze whether the dataset is suitable for factor analysis. The basic requirement is that a KMO value greater than 0.5 and a Bartlett test p-value less than 0.05 indicate that factor analysis can be performed.

As shown in Table 4, the KMO test coefficient is 0.654, and the Bartlett test result is 0.0040, which is less than 0.05. Therefore, the data structure of the research variables is reasonable, and the next step is to extract the principal components from the sentiment proxy variables.

Table 5. Total Variance Explained Table

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% Of Variance	Cumulative %	Total	% Of Variance	Cumulative %	Total	% Of Variance	Cumulative %
1	4.718	47.178	47.178	4.718	47.178	47.178	3.472	34.722	34.722
2	2.136	21.359	68.537	2.136	21.359	68.537	3.016	30.162	64.883
3	1.665	16.653	85.191	1.665	16.653	85.191	2.031	20.307	85.191
4	.527	5.272	90.463						
5	.406	4.059	94.522						
6	.307	3.069	97.591						
7	.085	.848	98.438						
8	.073	.734	99.172						
9	.052	.522	99.694						
10	.031	.306	100.000						

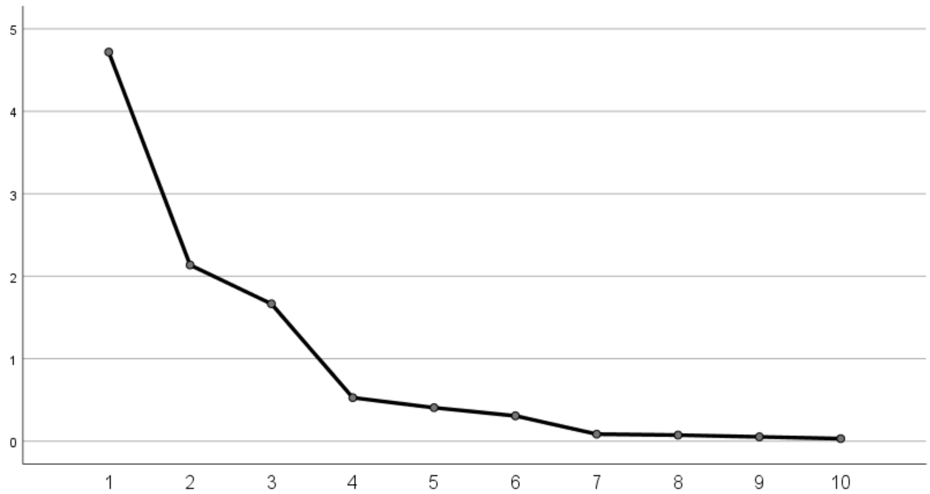

Fig. 1 Scree Plot of Principal Components

This ensures that the selected variables are sufficiently correlated and suitable for further analysis, allowing for a more accurate construction of the comprehensive investor sentiment index through principal component analysis. The KMO and Bartlett tests confirm that the dataset can provide meaningful insights into the relationships among the variables related to investor sentiment.

Based on the factor analysis of the ten variables mentioned above, a total variance explanation table can be obtained. According to the results in Table 5, the principal components used to construct the investor sentiment index can be determined. From the total variance explanation table in Table 5, it is clear that among the ten principal components, the first three principal components have eigenvalues greater than 1, specifically Principal Component 1, Principal Component 2, and Principal Component 3. Additionally, the cumulative variance contribution rate of these three principal components is 85.191% (which is greater than 80%).

Moreover, based on the slope of the scree plot in Figure 1, the curves in the first three segments almost cover most of the content. Therefore, the first three principal components can effectively explain the sentiment indicators that this paper aims to describe. On this analytical basis, this study ultimately selects Principal Component 1, Principal Component 2, and Principal Component 3 to construct the SENT (Investor Sentiment Index).

Table 6. Component Score Coefficient Matrix

	1	2	3
CCI	-.018	-.043	.493
TURN	.255	-.026	-.088
NNA	.058	.218	.003
PE	-.122	.373	-.045
MMV	.306	-.130	.079
CCI_{t-1}	-.014	-.043	.492
$TURN_{t-1}$	.245	-.007	-.101
MMV_{t-1}	.299	-.110	.068
NNA_{t-1}	.050	.224	-.006
PE_{t-1}	-.131	.381	-.046

To construct the principal components from the matrix in Table 3-6, we can denote the three principal components as (S1), (S2), and (S3). The variables (X1), (X2), (X3), (X4), (X5), (X6), (X7), (X8), (X9), and (X10) represent CCI, TURN, NNA, PE, MMV, and others, respectively. The expressions for these components can be written as follows:

S1=-0.018X1+0.255X2+0.058X3-0.122X4+0.036X5-0.014X6+0.245X7+0.299X8

+0.05X9-0.131X10

S2=-0.043X1-0.026X2+0.218X3+0.373X4-0.130X5-0.043X6-0.007X7-0.11X8

+0.224X9+0.381X10

S3=0.493X1-0.088X2+0003X3-0.045X4+0.079X5+0.492X6-0.101X7+0.068X8

-0.006X9-0.046X10

The expression for SENT can be formulated based on the variance contribution rates of principal components 1, 2, and 3, as follows:

SENT=34.722/85.191*S1+30.162/85.191*S2+20.307/85.191*S3

Table 7. correlation coefficient matrix

		TURN	MMV	NNA	PE	CCI
SENT0	Pearson Correlation	.740**	.553**	.776**	.796**	.130
	Sig. (2-tailed)	.000	.000	.000	.000	.061
	Number of Cases	207	207	207	207	207
		$TURN_{t-1}$	MMV_{t-1}	NNA_{t-1}	PE_{t-1}	CCI_{t-1}
	Pearson Correlation	.754**	.575**	.753**	.785**	.138*
	Sig. (2-tailed)	.000	.000	.000	.000	.048
	Number of Cases	207	207	207	207	207

Based on the correlation coefficient matrix between the comprehensive investor sentiment index and the ten variables presented in Table 7, we select the proxy indicators with higher correlation from both the current and lagged periods to further construct the comprehensive investor sentiment index. Therefore, we choose the following as the final indicators for constructing the comprehensive investor sentiment index: the lagged consumer confidence index, the lagged trading volume, the lagged turnover rate, the current price-to-earnings (P/E) ratio, and the current number of new accounts opened.

Table 8. component score coefficient matrix

	component		
	1	2	3
NNA	.060	.465	.010
PE	-.229	.710	-.115
$TURN_{t-1}$	.519	-.052	-.182
MMV_{t-1}	.568	-.193	.151
CCI_{t-1}	-.025	-.092	.971

Continue to perform principal component analysis on the five indicators, selecting the first three principal components to construct a comprehensive investor sentiment index.

$SENT\_pre1=0.06*NNA-0.229*PE+0.519*TURN_{t-1}+0.568*MMV_{t-1}-0.025*CCI_{t-1}$

$SENT\_pre2=0.465*NNA+0.710*PE-0.052*TURN_{t-1}-0.193*MMV_{t-1}-0.092*CCI_{t-1}$

$SENT\_pre3=0.01*NNA-0.115*PE-0.182*TURN_{t-1}+0.151*MMV_{t-1}+0.9712*CCI_{t-1}$

The final comprehensive investor sentiment index, weighted based on the variance contribution rates, is:

$SENT=36.428/89.469*SENT\_pre1+32.161/89.469*SENT\_pre2+20.88/89.469*SENT\_pre3.$

3.4. Trend of the comprehensive investor sentiment index

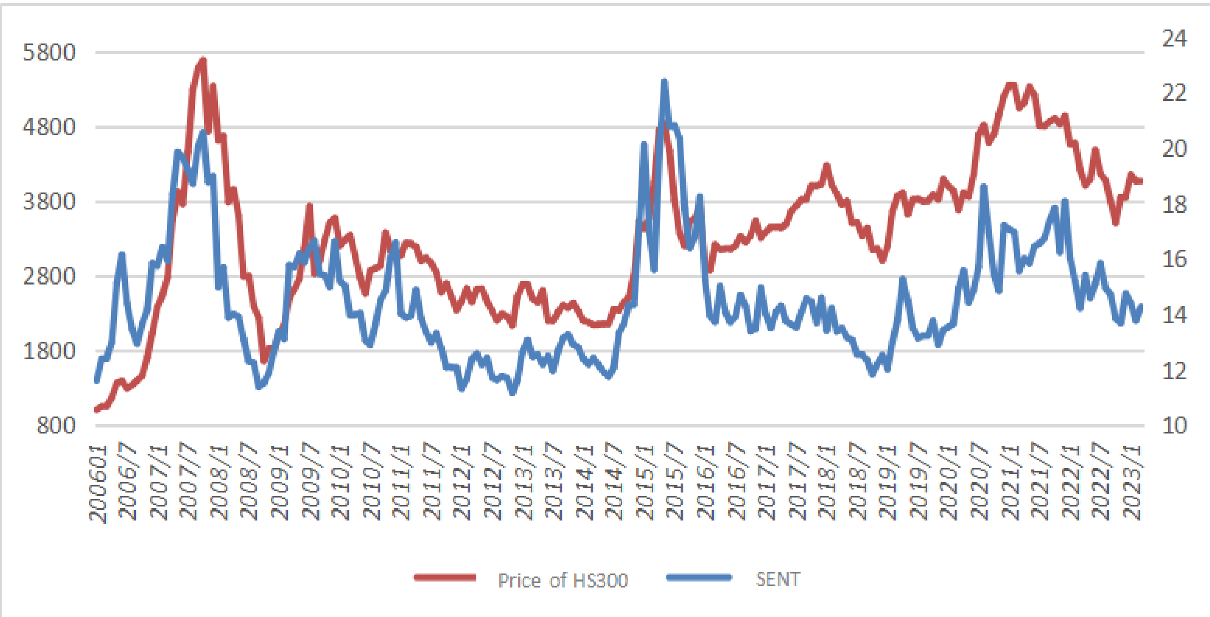

Fig. 2 Chart of the Closing Prices of the CSI 300 Index and the SENT Index

The CSI 300 Index, selected based on criteria of scale and liquidity, reflects the comprehensive price movements of representative stocks from the Shanghai and Shenzhen stock exchanges. This index is characterized by its stability and representativeness. Figure 2 presents a comparative line chart of the investor sentiment index and the CSI 300 Index. By visualizing the trends of the CSI 300 Index and the SENT Index, we can intuitively assess the effectiveness of the SENT Index. The chart demonstrates that the SENT Index largely aligns with the market movements of the CSI 300 across most intervals. This indicates that the investor sentiment index constructed in this study possesses a degree of rationality and explanatory power.

4. Empirical Test of the Impact of Investor Sentiment on Stock Returns

4.1. Single-factor model

The single-factor model for analyzing investor sentiment and stock returns is as follows:

$$R_{t+1} = c + \alpha SENT_t + \varepsilon_{t+1}$$

Where R represents the return rate of the CSI 300 Index, SENT represents the investor sentiment index constructed in this paper, c is the constant term, and ε is the residual.

In the previous section, this study analyzed the explanatory power of the investor sentiment index through correlation tests. To further confirm the validity of the constructed investor sentiment index, we employed a single-factor model regression for testing. The regression results are shown in Table 9:

Table 9. Single-factor model

	Variable	Regression Coefficient	t value	P value	R^2
Single-factor model	c	-15.72485	-3.594092	0.000	0.0687
	SENT	1.154945	3.878499	0.000	

The results of the single-factor model regression indicate a linear correlation between the returns of the CSI 300 index and investor sentiment, which is a significant positive correlation. This means that when the investor sentiment indicators increase, the returns of the CSI 300 index also rise in the same direction. Conversely, when the investor sentiment indicators decline, the returns of the CSI 300 index also decrease.

4.2. Robustness Test

Table 10. Single-factor model of the returns of each index

	Variable	Regression Coefficient	t value	P value	R^2
S50	c	-15.04160	-3.870708	0.000	0.0748
	SENT	1.073420	4.058506	0.000	
Z500	c	-14.80003	-3.438336	0.000	0.0645
	SENT	1.098309	3.748957	0.000	
SH	c	-14.75950	-4.255130	0.000	0.0885
	SENT	1.050473	4.449650	0.000	
SZ	c	-16.27104	-3.566899	0.000	0.0648
	SENT	1.167457	3.760247	0.000	

From the regression results, it is clear that there is a linear correlation between the Shanghai Stock Exchange 50 Index (SSE 50) return, the CSI 500 Index return, the Shanghai Composite Index (SSE Composite) return, and the Shenzhen Composite Index return with investor sentiment. This relationship is significantly positive, indicating that when the investor sentiment index increases, the returns of the SSE 50, CSI 500, SSE Composite, and Shenzhen Composite Index also rise correspondingly. Conversely, when the investor sentiment index decreases, the returns of these indices also decline. This indicates that the results are robust.

4.3. VAR Model

4.3.1. Stationarity Test

Conducting a unit root test is a prerequisite for establishing the VAR model. Therefore, it is essential to first test the stationarity of the variables included in the model. This study employs the Augmented Dickey-Fuller (ADF) test to examine whether the four financial time series—SENT, the CSI 300 Index, the CSI 500 Index, and the SSE 50 Index—are stationary, in order to avoid the issue of spurious regression. The test results are as follows:

Table 11. ADF TEST

Variable	ADF value	5% critical value	1% critical value	conclusion
SENT	-3.526386	-2.875752	-3.462901	stable
H300	-12.74721	-3.139664	-3.431896	stable
Z500	-12.35613	-3.431896	-4.003449	stable
S50	-12.97067	-3.431896	-4.003449	stable

In Table 11, H300, Z500, and S50 represent the log returns of the CSI 300, CSI 500 Index, and SSE 50, respectively. Based on the ADF values from the test results, we can conclude that SENT, H300, Z500, and S50 are all stationary.

4.3.2. Granger Causality Test

After confirming that the four financial time series under study are stationary, we proceed to conduct the Granger causality test for SENT, H300, Z500, and S50.

Table 12. Results of the Granger causality test between SENT and three types of representative stock market returns

Variable	F statistic	P value
SENT does not Granger-cause H300.	7.43696	0.0008
H300 does not Granger-cause SENT	4.04353	0.0190
SENT does not Granger-cause Z500	9.78919	0.0000
Z500 does not Granger-cause SENT	6.93941	0.0012
SENT does not Granger-cause S50	10.5647	0.0000
S50 does not Granger-cause SENT	7.11833	0.0010

From Table 12, we can see that the P-values for the listed null hypotheses are 0.0008, 0.0190, 0.0000, 0.0012, 0.0000, and 0.0010, all of which are significantly less than 0.05. Therefore, we reject the null hypotheses at the 5% significance level. This indicates that there is a Granger causality relationship between SENT and H300, Z500, and S50. We can now use these variables to construct the VAR model.

4.3.3. Lag Order Selection

Next, we determine the lag order required for the model, with the test results shown in Table 13. It can be observed that at the 5th lag, the LR, FPE, AIC, SC, and HQ criteria are all marked with an asterisk. Therefore, after comprehensive consideration, this study ultimately selects a maximum lag order of 2 for the model.

Table 13. Lag Order Selection Table for the VAR Model of SENT

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-1187.11	NA	419.7404	11.71539	11.74803	11.7286
1	-1021.25	326.823	85.1943	10.12068	10.21861	10.1603
2	-1009.89	22.16957*	79.23289*	10.04813*	10.21134*	10.11415*
3	-1006.07	7.358253	79.38369	10.04999	10.27849	10.14243

4.3.4. VAR Model Stability Test

Next, the article further examines the stationarity of the model. According to the distribution of AR roots, we can generally determine whether the model is stationary or non-stationary. Therefore, using the software to test the constructed model, we obtained the results shown in Figure 3.

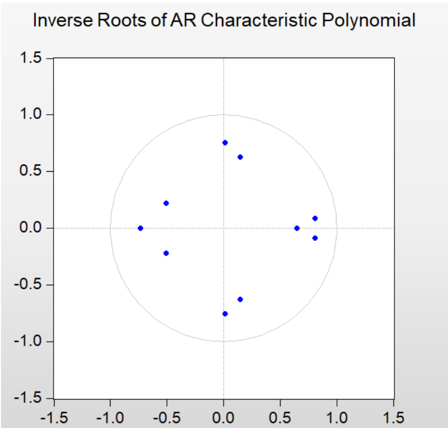

Fig. 3 Inverse Roots of AR Characteristic Polynomial Distribution Plot for Investor Sentiment and Stock Market Returns

According to the results shown in Figure 3, the AR roots of the VAR model are all located within the unit circle. Furthermore, based on the empirical results, the absolute values of the reciprocals of the characteristic polynomial roots of this model are all less than 1. Therefore, the model is stable, allowing for the continuation of subsequent impulse response analysis.

4.3.5. VAR Model Stability Test

(1) Impulse Response Analysis

When market conditions change, resulting in fluctuations in investor sentiment, such changes should affect not only the investor sentiment variable itself but also induce different impacts on the returns of various stock indices. Impulse response analysis can illustrate how a shock to investor sentiment, combined with disturbance terms, influences the current and lagged values of stock index returns that reflect price fluctuations.

In this study, impulse response analysis is employed to depict the specific effects of these fluctuations, providing a more intuitive understanding of how investor sentiment influences stock index returns and, in turn, how these returns affect investor sentiment.

The results of the impulse response analysis between the investor sentiment index (SENT) and the stock index returns for H300, Z500, and S50 are shown in Figure 4, respectively.

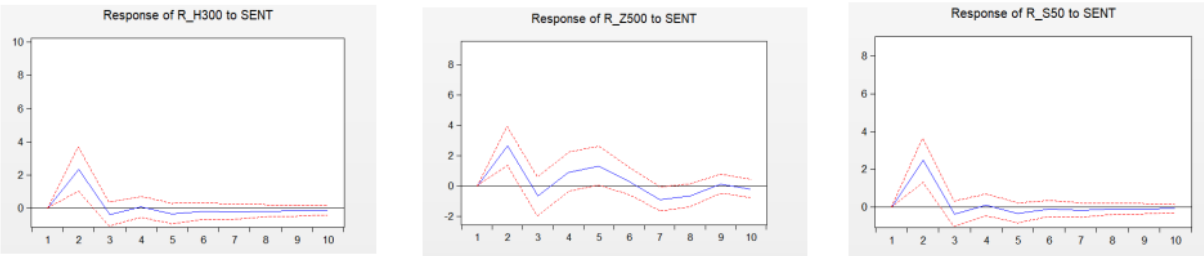

Fig. 4 Impulse Response Function Plot of SENT to CSI 300. CSI 500. SSE 50 Index Returns

Based on Figure 4, it can be observed that when the three stock indices experience shocks to investor sentiment, the H300, Z500, and S50 index returns do not respond immediately in the first period. This indicates that the impact of SENT on stock index returns exhibits a certain lag effect.

Both the H300, Z500, and S50 index returns show a significant positive response in the second period. This suggests that when investor sentiment rises, stock index returns quickly experience a positive impact in the short term. However, by the third period, the response of investor sentiment on stock index returns rapidly diminishes to negative values. This implies that the positive effect of investor sentiment on stock index returns is very short-lived. The initial rise in stock returns driven by increased sentiment may not be sustainable and could even lead to selling behavior by investors.

During the subsequent fourth to tenth periods, the stock index returns exhibit continuous fluctuations, oscillating up and down. By the tenth period, they essentially converge to zero. Therefore, it can be concluded that a shock to investor sentiment causes stock prices to first rise, then fall, and eventually stabilize.

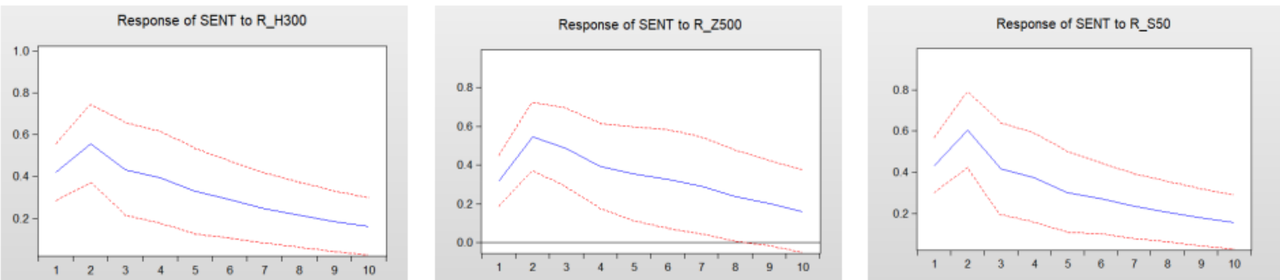

Fig. 5 Impulse Response Function Plot of CSI 300. CSI 500. SSE 50 to SENT Index Returns

As shown in Figure 5, when the H300, Z500, and S50 index returns experience shocks, SENT responds immediately in the first period. This indicates that stock index returns exhibit a significant positive reaction to SENT, meaning that when stock index returns rise, investor sentiment (SENT) quickly reflects a positive impact in the short term.

(2) Variance Decomposition Analysis

Table 13 illustrates the variance decomposition of investor sentiment, clearly showing the contribution of H300, Z500, and S50 index returns to SENT. Specifically, investor sentiment experiences the largest structural shock to itself, reaching 100% in the first period and gradually decreasing in the subsequent periods.Meanwhile, H300, Z500, and S50 index returns also exert some influence on investor sentiment. In the first period, the contribution of these index returns to investor sentiment is zero, indicating that the stock index returns do not contribute to SENT at that time, highlighting a certain lag in their impact on investor sentiment. By the tenth period, the contribution rates of the three indices return to investor sentiment have reached 1.648, 0.0609, and 1.415, respectively.

Table 14. Variance Decomposition Table of Investor Sentiment and Stock Returns

Period	S.E.	SENT	R_H300	R_Z500	R_S50
1	1.021114	100.0000	0.000000	0.000000	0.000000
2	1.403891	98.27578	1.724222	0.246615	2.359174
3	1.642951	98.44073	1.559275	0.160924	1.825341
4	1.797490	98.34805	1.651949	0.167440	1.731214
5	1.905317	98.36504	1.634962	0.131999	1.594332
6	1.981587	98.35384	1.646158	0.117062	1.535349
7	2.036819	98.35505	1.644947	0.097830	1.485057
8	2.077075	98.35283	1.647166	0.084085	1.454909
9	2.106672	98.35246	1.647541	0.071297	1.431502
10	2.128512	98.35175	1.648248	0.060850	1.415317

In summary, the H300 index returns have the greatest impact on investor sentiment, followed by the S50 index returns, while the Z500 index returns exert the least influence on investor sentiment. This indicates that when the stock market experiences fluctuations, investors first pay attention to the most representative large-cap stocks (H300 index), as significant movements in the market tend to affect the performance of most stocks.

Secondly, investors focus on the S50 index, which represents leading stocks in the market. The performance of these blue-chip stocks is often closely monitored, as they can influence the overall market trend. When the S50 index shows an upward trend, investor sentiment is generally optimistic, leading them to believe the market is performing well. Conversely, when the S50 index declines, investor sentiment may turn pessimistic, prompting caution regarding market movements.

In contrast, the Z500 index includes a larger number of smaller companies, which have a relatively minor impact on investor sentiment. Consequently, the number of investors focusing on the Z500 index is lower compared to the first two indices. This suggests that investors are more attuned to the movements of major indices that better reflect market performance and trends.

5. Conclusion and Limitations

This study begins with the definition of investor sentiment and employs ten variables to conduct principal component analysis, thereby constructing an investor sentiment index. The credibility of this index is verified through empirical research on the relationship between the investor sentiment index and the H300 index returns. The findings indicate that there is a linear and significantly positive correlation between the H300 index returns and investor sentiment. Specifically, when the investor sentiment index strengthens, the H300 index returns also increase; conversely, when the investor sentiment index declines, the H300 index returns also decrease.

Similarly, there are linear and significantly positive correlations between investor sentiment and the returns of the S50, Z500, SSE Composite Index, and SZSE Component Index. This means that as the investor sentiment index rises, the returns of these indices also tend to rise, and vice versa.

In studying the relationship between SENT and stock returns, this paper constructs a VAR model and uses impulse response analysis to reveal the mutual influence between SENT and the returns of the H300, Z500, and S50 indices. While SENT has a positive impact on the indices in the second period, this influence rapidly declines to negative in the third period. Over time, this impact fluctuates in the medium to long term, with decreasing amplitude, eventually converging to zero. In the context of China's stock market, while there is a significant presence of retail investors, their judgment is often immature, leading to herd behavior and speculative actions. Consequently, SENT can have a substantial short-term impact on the stock market, as participants may overtrade due to shifts in their subjective emotions. However, in the medium to long term, as investors gain more relevant market information, the influence of SENT on the stock market diminishes.

Predicting returns in practical applications is highly difficult and complex. Given the limited research capabilities and tight timelines in capital markets, this study has several shortcomings, which could be addressed in future research. First, this study did not conduct out-of-sample forecasting during the empirical research; subsequent studies should include rolling predictions for out-of-sample data. Second, this paper did not comprehensively consider all factors influencing the H300 index. Lastly, the variables considered in constructing the investor sentiment index were also limited.

Acknowledgements

The authors gratefully acknowledge the financial support from xxx funds.

References

[1] Cheng Qiyun, Sun Caixin, Zhang Xiaoxing, et al. Short-Term load forecasting model and method for power system based on complementation of neural network and fuzzy logic. Transactions of China Electrotechnical Society, 2004, 19(10): 53-58.

[2] Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.

[3] Amjady N. Short-term hourly load forecasting using time series modeling with peak load estimation capability. IEEE Transactions on Power Systems, 2001, 16(4): 798-805.

[4] Ma Kunlong. Short term distributed load forecasting method based on big data. Changsha: Hunan University, 2014.

[5] SHI Biao, LI Yu Xia, YU Xhua, YAN Wang. Short-term load forecasting based on modified particle swarm optimizer and fuzzy neural network model. Systems Engineering-Theory and Practice, 2010, 30(1): 158-160.

[6] Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.

[7] Amjady N. Short-term hourly load forecasting using time series modeling with peak load estimation capability. IEEE Transactions on Power Systems, 2001, 16(4): 798-805.

[8] Ma Kunlong. Short term distributed load forecasting method based on big data. Changsha: Hunan University, 2014.

[9] SHI Biao, LI Yu Xia, YU Xhua, YAN Wang. Short-term load forecasting based on modified particle swarm optimizer and fuzzy neural network model. Systems Engineering-Theory and Practice, 2010, 30(1): 158-160.

[10] Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.

[11] Amjady N. Short-term hourly load forecasting using time series modeling with peak load estimation capability. IEEE Transactions on Power Systems, 2001, 16(4): 798-805.

[12] Ma Kunlong. Short term distributed load forecasting method based on big data. Changsha: Hunan University, 2014.

[13] SHI Biao, LI Yu Xia, YU Xhua, YAN Wang. Short-term load forecasting based on modified particle swarm optimizer and fuzzy neural network model. Systems Engineering-Theory and Practice, 2010, 30(1): 158-160.