

SENTIMENT AND INDIVIDUAL STOCK PERFORMANCE: EVIDENCE FROM CHINA

Dr. Min Bai

Waikato Management School
University of Waikato, New Zealand
ORCID: <https://orcid.org/0000-0003-1249-7344>
Email: mbai@waikato.ac.nz

Prof. Yafeng Qin

School of Economics and Finance,
Massey University, New Zealand
ORCID: <https://orcid.org/0000-0002-8097-095X>
Email: y.qin@massey.ac.nz

Received: 2022-12-06

Accepted: 2023-01-25

Published online: 2023-01-31

Abstract

This paper investigates the effect of investor investment sentiment on individual stock returns in China. We find that investment sentiment is positively associated with stock performance contemporaneously. The Granger Causality test shows that investor sentiment is a driving force of stock price movement but not the other way around. Our constructed VAR model further suggests that the change in investor sentiment in the lagged period does not significantly affect the current stock return. In addition, both the Impulse Response and the Variance Decomposition analysis provide evidence that the stock price will increase right after a positive sentiment shock, and then follow a price drop until an intrinsic value comes up. Overall, the investor sentiment factor has a strong explanatory power to explain the variation in stock return.

Keywords: Investor Sentiment, Stock Return, Chinese Market, Principal Component Analysis, VAR Model.

JEL Classification: G12; G14; G15.

1. Introduction

Markowitz's Portfolio Theory (1952) is the pioneer of the traditional financial theory that suggests investors are rational and the investor's sentiment does not affect asset prices. Supporting this view, Efficient Market Theory (EMH), Fama (1970) also argues that investors in the market are rational and can make the rational investment to reflect the latest information. In the real world, however, the prices of most financial assets are not consistent with the theoretical predictions. Specifically, there are many anomalies that cannot be explained by the traditional financial theories such as the Calendar Effect, French (1980) and Gibbons and Hess (1981), the Equity Premium Puzzle, Rajnish and Prescott (1985), and the size effect, Banz (1981).

Thereby, by analysing the empirical results of cognitive psychology, emotional psychology and social psychology, a new strand of research has extended the original theories and developed new models to explain the investor's decisions. Behavioral finance suggests that investors' rational is limited when making investment decisions. Thus, they can only achieve limited arbitrage. From the perspective of investors, behavior literature shows that the price of financial assets is largely determined by investors' behavior and psychology on top of market factors. This paper focuses on the sentiment factor that has gradually become one of the main research issues in behavior finance.

Since the establishment of the Shanghai and Shenzhen stock exchanges, the Chinese stock market has experienced a number of abnormal fluctuations. From June 2005 to October 2007, the Shanghai Composite Index soared all the way from 1,060 to 6,100, an astonishing 500% increase in total. However, in the following year, the Shanghai Composite Index dropped more than 1,800 points, and the market value evaporated by 73%. The Shanghai Composite Index raised by nearly 160% from 2,000 level in May 2014 to 5,100 level in May 2015, and then sharply turned into a bearish market until February 2016. The indexes dropped more than 2,600, a near 50% decrease in just eight months. Due to the large and unforeseen price drops in the Chinese stock market, we argue that investor sentiment may be the driving force behind the drops. According to WIND and Guo Jin statistics reports, by the end of the third quarter of 2018, individual investors accounted for 30.3% of the stock market value, while institutional investors accounted for 12.6%, showing that the proportion of specialized institutional investors is far less than individual investors. It implies that institutional investors are no longer playing a leading role in the Chinese stock market. In this case, individual investors are easy to be affected by various aspects of irrational behavior such as the market sentiment effect and the herding effect. This study investigates the relationship between the investor's sentiment factor and the stock market performance as well as whether there is a causality effect between the sentiment and the stock returns. Consequently, this study will shed the light on exploring the relationship between investor sentiment and stock market returns, especially by providing empirical evidence to the behavioral finance theory and the sentiment effect.

The remainder of this paper is organized as follows. Section 2 presents the literature review, sentiment index construction and hypotheses development. Section 3 describes the data, sample selection and methodology. Section 4 reports the empirical results and findings and Section 5 concludes.

2. Literature Review, Index Construction, and Hypotheses Development

2.1. Literature Review

The EMH, efficient market hypothesis, Fama (1970) is one of the most important theories in the field of traditional finance. The theory predicts that asset prices reflect all the relevant information, and future asset price changes cannot be predicted by the past price information. Moreover, all investors are rational when making their investment decisions and realize unlimited arbitrage. The investor sentiment model (BSV model) proposed by Barberis, Shleifer, and Vishny (1998) believes that investors will be affected by conservatism and representative bias when facing the new financial reports of listed companies, which will lead to some common anomalies. Daniel, Hirshleifer, and Subrahmanyam (1998, 2001) focus on the cognitive bias of investors when dealing with private information. They suppose that investors have done a lot of in-depth research on a listed company. In doing so, investors tend to be overconfident in their own analysis results. If the results of the analysis show that the company's basic orientation is good, they will buy the company's shares on a large scale; in addition, in the process of confirmation bias, they will only focus on the public information consistent with their own analysis results in the short term, and ignore the information with different opinions, which will lead to such phenomena as momentum and PEAD. In addition, Charles (2009) examines the causes of the financial crisis and subsequent economic recession and attributed the asset price bubble and subsequent collapse to investor overconfidence in a low volatility moderate macroeconomic environment.

Empirically, since the 1980s, shreds of evidence show that the "market anomalies" findings deviate from the traditional efficient market hypothesis. E.g. The "Calendar effect" is one of the earliest observed market anomalies. The effect suggests that there are systematic differences in financial asset returns over different periods of time. The calendar effect mainly includes the weekly effect and the monthly effect, the seasonal effect and the holiday effect, which respectively refer to the abnormal second moments of abnormal returns and other abnormal higher moments related to the week and holiday of the season and month in the financial market. This cyclical anomaly is contrary to the efficient market hypothesis, because asset returns are no longer random walks, but are based on certain predictability in a specific calendar period, French (1980) and Gibbons and Hess (1981).

Due to the investors' irrational behaviors, a behavior finance factor – Sentiment is getting increasing attention recently. Investors usually cannot process information correctly, so they cannot correctly infer the probability of price trends. Even if the probability of a future price trend is given, investors usually make inconsistent or suboptimal decisions. Financial market participants will have emotional fluctuations and investor sentiment is uncertain, which makes it difficult to measure investor sentiment.

The change in investor sentiment will affect their prediction of the future trend of the stock market. The wrong prediction leads to wrong decision-making. The wrong decision-making of a large number of investors at the same time will have a significant impact on the capital market. Therefore, investor sentiment is very important in the stock market.

Based on the limited rationality of investors, early literature defined investor sentiment as the investors' expectation deviation of stock price. Zweig (1973) studies the deviation between investor sentiment and expected returns and defines the difference between the estimated value and the true value of securities as a proxy of investor sentiment. After then, some researchers found that there are many noise traders in the market, and their irrational behavior often makes investors in the market adjust their original investment decisions. In other words, the irrational behavior of noise traders may be the cause of investor sentiment. Black (1986) investigates the irrational behavior of noise traders and confirmed that the investors' expectation bias was mainly caused by noise traders. However, none of the above research explained what investor sentiment is. Based on previous studies, Lee et al. (1991) for the first time elaborate in detail what investor sentiment is, and argue investor sentiment is a price expectation that cannot be explained by fundamental factors. Shleifer et al. (1997) suggest that when people make investment decisions, due to the existence of bounded rationality, they can accurately judge when to achieve the expected utility. Therefore, this common misjudgment of investors is defined as a proxy of investor sentiment. Further research, Shleifer (2000) shows that investors have different expectations for future returns of investable assets. Therefore, this systematic expectation bias is further defined as an indicator of investor sentiment.

More studies introduce investor sentiment into the asset pricing model of traditional finance. One stream of research argues, investors will have their own psychological price for assets which they believe in. Therefore, investor sentiment is defined as a pricing belief or confidence. Specifically, Barberis (1998) thinks that the fluctuation of a stock price is mainly caused by noise traders in the market. He conducts an empirical test and confirms that the pricing belief of noise traders is an important factor in the change of stock price and defines this belief as investor sentiment. De long et al. (1990) thinks that investors mainly focus on cash flow and risk when they participate in market transactions. However, when the current facts cannot confirm the trend of the stock market in the future, the confidence of investors will play an important role in their views on the future market expectations. Therefore, he defines the confidence of investors as investors' sentiment. In addition, Baker and Stein (2004) document that investors cannot be completely rational so they can not accurately evaluate the price of securities. Investor sentiment is defined as the inborn speculative tendency of investors. Based on the extant literature on investor sentiment, we adopt the following investor sentiment measures in our study.

2.2 Measurement of Investor Sentiment

The measurement of investor sentiment is one of the key issues in behavioral finance. There are two main methods to measure investor sentiment: the single index method and the comprehensive index method. The single index method is further divided into the direct index method and the indirect index method. The direct index is obtained through a questionnaire survey or interview. The questionnaire survey objects include individual investors and institutional investors. The indirect index refers to the selection of financial transaction data that can reflect investor sentiment in the financial market. The comprehensive index represents the comprehensive index constructed by using principal component analysis, partial least squares regression, and other measurement methods.

Single Index Method

The association index and Wall Street analyst sentiment index are used as a single direct indicator of investor sentiment in previous literature.

The investor intelligence index (II Index) is mainly used to describe investor sentiment by collecting the number of professional investors who have different opinions on the future stock market trend and calculating the percentage difference between the number of bullish and bearish investors. The index is compiled by a company called Chartcraft since 1964.

The association of individual investors index (AAII Index) is used to represent the investor sentiment in the U.S. market. The American Association of individual investors conducts a questionnaire survey of its members on a monthly basis. The AAII index is obtained by recording the members' forecasts of the stock market trend in the next half year and calculating the proportion of each trend. This index is monthly data and reflects the sentiment of individual investors.

The Wall Street analyst sentiment index is mainly used to replace the sentiment of institutional investors because the index is obtained by calculating the proportion of stocks in the stock portfolio recommended by seller analysts. A number of studies have done a lot of research on this index to represent investor sentiment, but the results are still inconsistent.

Indirect indicators of the index - Indirect indicators of investor sentiment are trading data in the stock market to represent sentiment. Most of these indicators reflect market activity. Some are often used as indirect indicators: Closed-end Fund Discount, IPO Number, IPO First Day Return Rate, Turnover Rate, NEW, VOL, P/E, Margin Trading, Short Selling, etc.

The discount rate of closed-end fund (DCEF) refers to the ratio of the difference between the net asset value and the price of the fund to the net asset value. This index

reflects the deviation degree between the price and the value of the closed-end fund. It is the most often used index to represent investor sentiment. Lee et al. (1991) discover that the discount rate of closed-end funds significantly affected the small-cap stocks, which could reflect the sentiment of individual investors. Based on their theory, Brown and Cliff (2005) draw the same conclusion as Lee et al. (1991).

The number of IPOs and the First Day's Return of IPOs are also used to describe investor sentiment. Based on IPO trading volume, Brown and Cliff (2005) show that there is a relation between stock returns and IPO-related indicators. Ljungqvist et al. (2006) document that when market sentiment is high; IPO is more likely to be issued in the market which makes IPO-related indicators a proxy variable of sentiment.

Turnover rate (TURN) refers to the proportion of stocks sold and bought by investors in the market, which reflects the investors' investment behavior of predicting the future price changes of the stock market according to the market information. The turnover rate is an indirect indicator used as the most important indicator of investor sentiment. Baker and Stein (2004) find that the turnover rate essentially reflects the stock liquidity in the stock market, the stock liquidity is very concern for both investors and speculators and the turnover rate will certainly reflect the corresponding sentiment. Baker and Wurgler (2006) find the same results as that of Baker and Stein (2004), which confirms once again that turnover rate can be used as a proxy index of sentiment.

The number of new investor accounts (NEW) is rarely used to represent investor sentiment, mainly due to the capital markets are getting more and more mature, and the NIA index changes little, which cannot reflect the change in investor sentiment. Shiller (2005) believes that the formation of a bull market is consistent with the sharp increase in the number of people who directly participate in the stock market but not through institutions.

Trading volume (VOL): Trading volume can reflect the liquidity of the stock market. Generally, the larger the trading volume is, the higher the liquidity is. When the investor sentiment is high, the trading volume will be enlarged.

Market Price-to-Earnings Ratio (MP): The P/E ratio is equal to the ratio of stock price to annual earnings per share, which is the most used index for fundamental analysis. In addition to analyzing the profitability of the company, the index can also reflect the investors' expectations of the future return of the stock. Therefore, the higher the P / E ratio, the more confident the investors will be, which will lead to an upsurge in market performance and sentiment.

Margin Trading and Short Selling. Since the implementation of margin trading on 31st March 2010 in China by following Hong Kong's Short-Sales setting, Bai, et al., (2017) and Bai (2021), the scope of subject-matter securities of margin trading in

Shanghai and Shenzhen stock markets has been continuously expanded and the stock varieties available for trading have been continuously enriched and diversified. Margin trading can increase market liquidity through leverage, which has a great impact on the stock market. The higher the financing balance, the more optimistic the investor's sentiment is, and they are willing to increase leverage to buy stocks.

Comprehensive Index Method

The comprehensive index of investor sentiment is to add several single indicators through a mathematical method, and finally get a comprehensive index that can reflect the information of multiple indicators. BW index is the most accepted index among the comprehensive indicators. It is a comprehensive sentiment index obtained by Baker and Wurgler (2006) by selecting six single indicators. The empirical results show that the comprehensive index is more significant than the single index. Glushkov (2006) uses eight direct and objective single indicators including closed-end fund discount rate, II index, trading volume; bull-bear market ratio, IPO number and first-day return to construct a composite index. Baker et al. (2009) construct the global investor sentiment composite index (GSI) and selected single indicators: volatility premium, trading volume, IPO quantity. Using the principal component analysis method, the local sentiment index of each sample country was established, and then the local sentiment index of the sample country was synthesized by principal component analysis to construct the global Comprehensive index of emotion.

However, the BW index and some indexes based on BW still have many shortcomings: In the process of the index construction, only the first principal component is selected, and its variance interpretation rate is only about 50%, which leads to the loss of any information. Secondly, the BW index is an annual index, which can truly reflect the changes in investor sentiment in mature stock markets; nevertheless, in emerging markets like China, the market has only been established for a short time, the measurement of investor sentiment even takes a shorter time. Therefore, the annual data may not meet the requirements of empirical research.

To sum, the effectiveness of integrated investor sentiment is stronger, which can include more investor sentiment factors in the market, and the regression result is more significant. Therefore, we are constructing a new comprehensive sentiment index to capture not only the sentiment of institutional investors but also the sentiment of individual investors to facilitate the features of the Chinese stock markets. A comprehensive investor sentiment index is constructed by principal component analysis in this study.

2.3 The Impact of Investor Sentiment on Stock Market Returns

As for the research on the impact of investor sentiment on stock market returns, the main results are summarized in two aspects. The first is the overall effect of

investor sentiment on the stock market. The second is the cross-sectional effect of investor sentiment which is the impact of investor sentiment on the stock of a single company.

2.3.1 The Overall Effect of Investor Sentiment on Stock Market Returns

Baker et al. (2006) adopt the comprehensive investor sentiment index method and find that there is a common change relationship between investor sentiment and stock returns. Schneller et al. (2018) then use the empirical similarity approach to document that the sentiment of European investors significantly affects the volatility of European stock returns. Another interesting study about the factor that may influence investor sentiment and eventually affect the stock return was conducted by Drakos (2010); he finds that terrorist activities affect investor sentiment, terrorism led to a significant reduction in returns on the day of the attack.

Some studies further document that the relationship between investor sentiment and stock returns depends on time length, that is, the long-term and short-term effects are different. However, some literature documents inconsistent conclusions: Investor sentiment will affect stock market returns in reverse. Fisher et al. (2000) argue that there is a strong negative correlation between investor sentiment and the returns of the stock market.

2.3.2 Cross-Sectional Effect of Investor Sentiment on Market Stock Returns

The cross-sectional effect of investor sentiment on stock returns refers to the influence of investor sentiment on individual stock returns. Lee et al. (1991) look at the impact of investor sentiment on the stock portfolio return of NYSE-listed companies. The results show that investor sentiment has an impact on stock market returns. Dallin (2020) discusses the time-varying institutional investors' preference for lottery stocks and finds that when the investor's mood is low, institutional investors have extremely high trading profits on more active stocks. It is found that even if the company's financial situation is good, the company's name will cause investor sentiment to fluctuate.

2.4 Hypotheses Development

According to the existing investor sentiment indicators, this paper follows the principle of interpretability and the principle of continuity in determining the sentiment indicators. The principle of continuity refers to the suitability of selected indicators for long-term research. In addition, based on the background and characteristics of investor sentiment in China's stock market, this paper constructs an investor sentiment measurement index suitable for China's stock market and uses quantitative methods to evaluate the effectiveness of the investor sentiment index.

It is found that there is no consistent conclusion on the relationship between investor sentiment and stock returns in either overall effect or cross-section effect. Some studies have shown that investor sentiment has a positive impact on stock returns, while others have shown that investor sentiment has a negative or no effect on stock returns. The differences in these conclusions are also affected by the differences in investor sentiment indicators, research samples and data selection. Most of the research results on investor sentiment and stock returns support that investor sentiment has an impact on stock returns, but there are inconsistent conclusions on the predictive ability of investor sentiment. Most studies believe that investor sentiment is positively correlated with stock returns in the short term and negatively correlated with stock returns in the long run. However, some studies suggest that there is a negative correlation between investor sentiment and stock returns in the short term. Based on the mixed views, we propose the following hypothesis,

Hypothesis 1: There is a positive correlation relationship between investor sentiment and stock return.

Hypothesis 2: Investor sentiment's predictive and explanatory ability to stock return is significant but not the other way around in China.

3. Data and Methodology

The data used in this paper are downloaded from the WIND database and the PE is a dynamic P/E ratio obtained by adding the market value of the Shanghai and Shenzhen stock markets. Since China began to implement margin trading in March 2010, the data sample in this paper is from March 2010 to December 2019.

Principal component analysis (PCA) was proposed by Hotelling (1933). The main principle is to use mathematical transformation to reduce the dimension of multi-dimensional explanatory variables. According to the analysis results, we can see the contribution degree of each variable to the explained variable. We remove the variable with a very small contribution value and take the first few explanatory variables with an 85% contribution value. This makes the model simple and reduces the occurrence of collinearity.

VAR Regression analysis: For the empirical analysis of the influence of investor sentiment on the return rates of Chinese stock markets, this paper first uses the ADF test to see the data's stability. If the data is stable, then we use the VAR model, Granger Causality Test, Impulse Response and Variance Decomposition analysis to test the relation between investor sentiments and return rate respectively.

3.1 Selection of Proxy Variables of Investor Sentiment

This paper uses the seven indicators/proxies of investor sentiment: Sentiment proxies' construction:

Definition of the 7 sentiment proxies

DCEF	Discount Rate of Closed-End Fund	$DCEF_t = \frac{\sum_{i=1}^n [(P_{it} - NAV_{it}) \times N_i]}{\sum_{i=1}^n (N_i \times NAV_{it})}$
MP	Market P/E Ratio	Stock Price Per Share / Earnings Per Share
TURN	Turnover rate of A-share market	(Trading Volume of The Month * Average Trading Days Over Months) / (Total Number of Shares in Circulation * Trading Days of Current Month)
NEW	New Investor Accounts	Number of accounts at the end of this month – Number of accounts at the end of last month
MARGIN	Margin Trading	Number of Margin Trading
SHORT	Short Selling	Number of Short Selling
VOL	Trading Volume	Number of Shares Traded

Note: In the formula for calculating the Discount Rate of Closed-End Funds: n is the number of closed-end funds. P_i is the closing price of fund i on the last trading day of each month. NAV_{it} is the net unit value on the last trading day of each month. N_i is the share of fund i . The unit of NIA is ten thousand.

It can be seen from Table 1 that the seven standardized variables have passed the correlation test. From the correlation coefficient of each variable, the correlation coefficient of new investor number (NEW) with trading volume (VOL) and margin trading (MARGIN) is 0.749 and 0.659 respectively, which has a strong positive correlation. This shows that in a good market situation, an increase in the number of new investors lead to an increase in the trading volume of the two markets. It implies that the investors are positive in this situation.

TABLE 1
Correlation coefficient of standardized variables

	DCEF	NEW	MP	TURN	VOL	MARGIN	SHORT
DCEF	1						
NEW	-0.0670	1					
MP	-0.163*	0.696***	1				
TURN	-0.320***	0.777***	0.595***	1			
VOL	0.0840	0.749***	0.443***	0.854***	1		
MARGIN	0.343***	0.659***	0.410***	0.598***	0.840***	1	
SHORT	0.490***	0.222**	-0.198**	0.175*	0.541***	0.650***	1

Table 1 shows the correlations among indicators. Note: *, ** and *** indicate the significance level of 10%, 5% and 1%.

Based on the self-construction of the investor sentiment index, this paper further studies the relationship between investor sentiment and stock returns. The stock return rate is represented by RET and uses the Shanghai and Shenzhen 300 index. The monthly rate of return is calculated as $RET_t = P_t - P_{t-1}$. RET_t refers to the monthly return of the CSI 300 index. P_t and P_{t-1} refer to the closing price at the end of this month and the closing price at the end of last month.

3.1.1 Descriptive statistics of proxy variables of investor sentiment index:

In this study, the descriptive statistical analysis is carried out on the proxy scalar of seven sentiment indicators and the basic characteristics of each variable are qualitatively observed. The results are shown in Table 2.

TABLE 2
Descriptive Statistical Results of Investor Sentiments Indexes

Variables	Sample Volume	mean	sd	median	min	max	skewness	kurtosis
DCEF	118	-0.0826	0.0389	-0.0864	-0.184	-0.0213	-0.316	2.225
NEW	118	38.03	27.96	33.62	9.190	165.8	2.625	11.70
MP	118	23.65	5.807	22.69	14.69	41.67	0.570	2.526
TURN	118	256.0	121.8	224.9	103.6	767.7	1.996	7.283
VOL	118	3.320e+11	2.100e+11	2.930e+11	8.680e+10	1.020e+12	1.465	5.075
MARGIN	118	5939	4784	7621	0.0600	20729	0.411	2.851
SHORT	118	39.05	31.67	34.33	0.0200	137.4	1.026	3.996

Table 2 illustrates descriptive statistics of the seven sentiment indicators. DCEF is Discount rate of closed-end fund; MP is market p/e ratio; TURN is turnover rate of A-share market; New is new investor accounts; VOL is trading volume; MARGIN is margin trading; SHORT is short selling. (Definitions are shown in section 3.1)

It can be seen that the sample observation data in this paper are collected from the monthly data during the period of March 2010 to December 2019 which consists of 118 groups. All the data are real and effective. According to Table 2, there is little difference between the maximum and minimum value of PE and VOL between the two markets. While the difference between the two values of New Investors is large. This indicates that a large number of new investors will enter the stock market when the market is good, which can also partly reflect investor sentiment.

3.1.2 Constructing the comprehensive index of investor sentiment by principal component analysis

Due to the differences in index units and dimensions, the seven variables are standardized to eliminate the impact on data and results. Thereby, KMO and SMC tests

are carried out¹. As can be seen from Table 3, the value of the KMO test is 0.711, which meets the standard of greater than 0.6. This result indicates that the correlation between various indicators is strong, which is suitable for factor analysis and the seven proxy variables selected in this paper can be used for principal component analysis. It also indicates that the principal component analysis in this paper has a good effect on the data constraint of 7 variables.

TABLE 3
The KMO and SMC value of the variables

Variable	KMO	SMC
DCEF s	0.466	0.624
NEW s	0.890	0.743
MP s	0.656	0.689
TURN s	0.650	0.919
VOL s	0.701	0.933
MARGIN s	0.816	0.833
SHORT s	0.650	0.730
Overall	0.711	

Table 3 shows the KMO and SMC results. Kaiser Meyer Olkin's Sampling Adequacy Measure (KMO) is used to measure the strength of the correlation between variables, which is obtained by comparing the correlation coefficient and partial correlation coefficient of two variables.

In order to construct the composite investor sentiment index (SENT), this paper needs to obtain the eigen values, cumulative variance contribution rate and component matrix of seven sentiment indicators. They are shown in table 4. Specifically, the characteristic roots of the first two principal components are greater than 1. However, according to the extraction rules of principal components, the cumulative variance is not less than 85%, and the first three factors can be selected as the principal components. Furthermore, the cumulative variance interpretation rate of the first three components reaches 91.9%.

¹ Kaiser Meyer Olkin's Sampling Adequacy Measure (KMO) is used to measure the strength of the correlation between variables, which is obtained by comparing the correlation coefficient and partial correlation coefficient of two variables. KMO is between 0 and 1. The higher the KMO means the stronger the commonality of variables. If the partial correlation coefficient is high relative to the correlation coefficient then KMO is relatively low which indicates that principal component analysis cannot play a good role in data reduction. According to Kaiser (1974), the general criteria are as follows: 0.00-0.49 (unacceptable); 0.50-0.59 (miserable); 0.60-0.69, (mediocre); 0.70-0.79 (middling); 0.80-0.89 (meritorious); 0.90-1.00 (marvelous). SMC is the Square of the Complex correlation coefficient between one variable and all other variables which is the determinable coefficient of the complex regression equation. The higher SMC ratio indicates the stronger the linear relationship of variables and the stronger the commonness means the more suitable principal component analysis is.

TABLE 4
Contribution Rate of Characteristic Root and Cumulative Variance

Component	Eigenvalue	Proportion	Cumulative
1	3.834	0.548	0.548
2	1.880	0.269	0.816
3	0.721	0.103	0.919
4	0.213	0.0304	0.950
5	0.195	0.0278	0.978
6	0.119	0.0171	0.995
7	0.0380	0.00540	1

Table 4 shows the Contribution Rate of Characteristic Root and Cumulative Variance. According to the extraction rules of principal components that the cumulative variance is not less than 85%, the first three factors can be selected as the principal components. Furthermore, the cumulative variance interpretation rate of the first three components reaches 91.9%.

Figure 1 is a Scree plot, which can intuitively show the size of each characteristic root. From the figure, component 1-3 is relatively important, while the following components are not very important. Figure 1 is a Scree plot showing the size of each characteristic root intuitively.

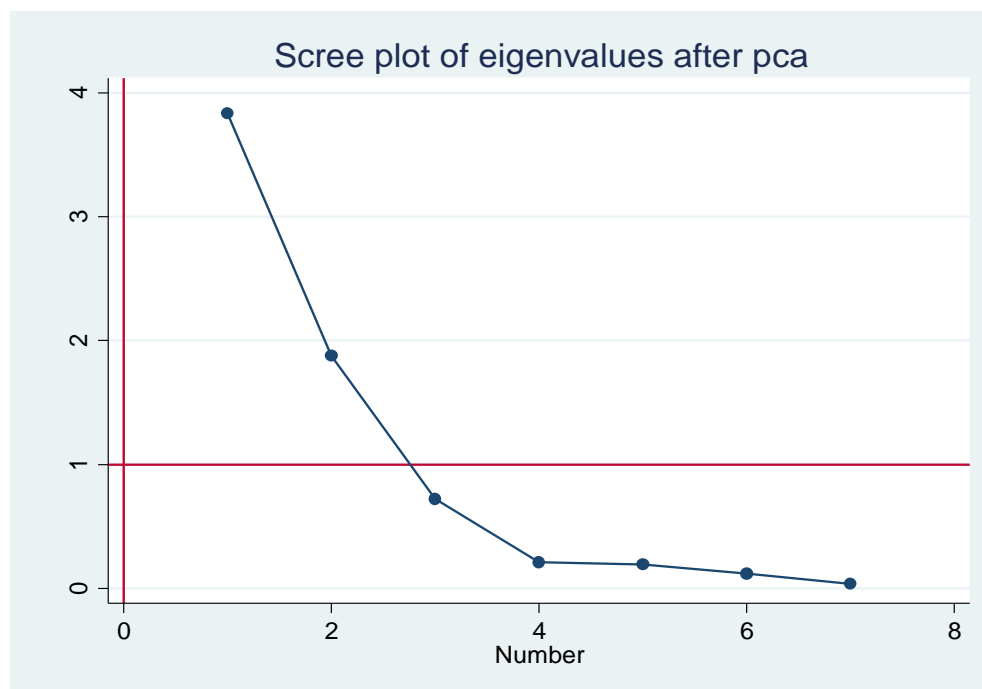


Figure 1. Scree plot of eigenvalues after PCA

The following figure shows the linear combination (score) of each principal component corresponding to each variable. It can be seen from Table 5 that the first principal component has the maximum factor load on the VOL-s index, the second principal component has the largest factor load on the DCEF-s index and the third principal component has a larger factor load on DCEF-s and MP-s indexes. In this paper, F1, F2 and F3 are used to represent the three principal components.

TABLE 5
Eigenvector table of principal components

Variable	Comp1	Comp2	Comp3
DCEF_s	0.0305	0.6028	0.607
NEW_s	0.4517	-0.1641	0.1563
MP_s	0.3258	-0.3805	0.5965
TURN_s	0.4438	-0.2539	-0.2964
VOL_s	0.4834	0.0796	-0.2167
MARGIN_s	0.4478	0.2605	0.0734
SHORT_s	0.2402	0.5712	-0.3333

Table 5 shows the linear combination (score) of each principal component corresponding to each variable after model's calculation.

The three principal components are respectively expressed by F_i according to their order. The calculation formula of principal components is as follows :

$$F1 = 0.0305 * DCEF_s + 0.4517 * NEW_s + 0.3258 * MP_s + 0.4438 * TURN_s + 0.4478 * MARGIN_s + 0.4834 * VOL_s + 0.2402 * SHORT_s$$

$$F2 = 0.6028 * DCEF_s - 0.1641 * NEW_s - 0.3805 * MP_s - 0.2539 * TURN_s + 0.0796 * VOL_s + 0.2605 * MARGIN_s + 0.5712 * SHORT_s$$

$$F3 = 0.607 * DCEF_s + 0.1563 * NEW_s + 0.5965 * MP_s - 0.2964 * TURN_s - 0.2167 * VOL_s + 0.0734 * MARGIN_s - 0.3333 * SHORT_s$$

After getting the formula of the principal component and weight by combining the variance contribution rate of Table 4, the initial investor sentiment is shown below:

$$SENT = 0.5477 * F1 / 0.9193 + 0.2686 * F2 / 0.9193 + 0.1031 * F3 / 0.9193$$

4. The Relationship between Investor Sentiment and Stock Returns

4.1 Correlation Test of Variables

To find out the relationship between investor sentiment and the overall return of Chinese stocks, it is first necessary to determine the correlation between the index RET used in this article to replace the overall return of Chinese stocks and the comprehensive sentiment index SENT. The results are shown in Table 6. Through the correlation test of investor sentiment and CSI300 index, it is found that the correlation coefficient of RET and SENT is 0.291 at the significance level of 5%, which indicates that RET and SENT have a positive and significant correlation.

TABLE 6
Pearson correlation coefficient analysis table

	RET	SENT
RET	1.000	
SENT	0.219**	1.000

Note: *, ** and *** indicate the significance level of 10%, 5% and 1%.

4.1.2 Stability Test of Time Series

Because the data selected in this paper are time series, the stationarity of the data needs to be verified. To optimize the stationarity, there is a risk of "spurious regression" when selecting a non-stationary time series for regression analysis. This paper uses the ADF test (Augmented Dickey-Fuller Test) to test the stability of investor sentiment and stock returns. The test results are shown in Table 7.

TABLE 7
Results of Unit Root Test

var	test statistic	5% critical value	P value	conclusion
RET	-1.805	-2.889	0.3781	not reject
SENT	-2.186	-2.889	0.2113	not reject
D(RET)	-15.503	-2.889	0.000	reject
D(SENT)	-10.11	-2.889	0.000	reject

Note: D (RET) D (SENT) refers to the difference variable of RET and SENT

It can be seen from Table 7 that after the ADF test, the t values of RET and SENT are -1.805 and 2.186 respectively, and the corresponding P values are 0.3781 and 0.2113, which are greater than 0.05. This result means that under the significance level of 5%, there are unit roots in these sequences indicating that the sequence is unstable. After the first-order difference, the corresponding p values of these tests are far less than 0.05, which shows that RET and SENT sequence after the first-order difference is in a stable state at the significance level of 5%.

4.2 VAR Regression Analysis

Table 8 shows the results of the VAR model. Column 1 is the regression result with RET being y, column 2 is the regression result with SENT being y, where RET (-1) refers to the first order delayed RET, and RET (-2) refers to the second-order delayed RET. The P values of ARCH-LM are greater than 0.05, indicating that there is no arch effect and the VAR model can be used for analysis.

TABLE 8
Results of VAR Regression

	(1)	(2)
	M1	M2
VARIABLES	RET	SENT
RET (-1)	0.083 (0.69)	0.0345*** (4.25)
RET (-2)	-0.0001 (-0.00)	0.0082 (0.98)
RET (-3)	0.029 (0.29)	0.0085 (1.27)
SENT (-1)	0.8267 (0.49)	0.5902*** (5.18)
SENT (-2)	-2.681 (-1.39)	-0.0891 (-0.68)
SENT (-3)	1.4915 (0.92)	0.4086*** (3.74)
Constant	0.5274 (0.82)	0.0187 (0.43)
Observations	115	115
R2	0.0359	0.8658
F	0.6701	116.177***
P	0.674	0
ARCH-LM	0.358	1.558
P	0.5497	0.2119

Table 8 shows the results of VAR model. Column 1 is the regression result with RET being y, column 2 is the regression result with SENT being y, where RET (-1) refers to the first order delayed RET, and RET (-2) refers to the second-order delayed RET. The P values of ARCH-LM are greater than 0.05, indicating that there is no arch effect and VAR model can be used for analysis. Note: t-statistics in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The coefficient of SENT (-1) is 0.8267, which means that the return of the current period will increase by 0.8267 units for every increase in investor sentiment of one-period lag. The coefficient of SENT (-2) is -2.681, which means that the return of the current period will decrease by 2.681 units for every increase in investor sentiment

in the second period of lag. The coefficient of SENT (-3) 1.4915 means that the yield of the current period will increase by 1.4915 units for every increase in investor sentiment in three periods. From the regression of column 2, we can see that RET (-1) is positively significant at the 1% level, which indicates that the increase in return rate in the lag period will significantly increase the investor sentiment in the current period. Specifically, the return rate of the lag period increases by 1 unit, and the current investor sentiment increases by 0.0345 units. Furthermore, SENT (-1) is positively significant at the 1% level, which indicates that the increase of investor sentiment in the lag period will significantly increase the investor sentiment in the current period. Specifically, the investor sentiment in the lag period will increase by 1 unit, and the investor sentiment in the current period will increase by 0.5902 units. SENT (-3) is also positively significant at the 1% level, indicating that the increase of investor sentiment in the lag period will significantly increase the investor sentiment in the current period. It indicates that the investor sentiment in the three lag periods will increase by 1 unit, and the investor sentiment in the current period will increase by 0.4086 units. The theoretical model of this paper can be written as:

$$R_t = \alpha_0 + \sum_{i=1}^n \alpha_i R_{t-i} + \sum_{i=1}^n \beta_i SENT_{t-i} + \varepsilon_{1t}$$

$$SENT_t = b_0 + \sum_{i=1}^n \gamma_i R_{t-i} + \sum_{i=1}^n \delta_i SENT_{t-i} + \varepsilon_{2t}$$

4.3 VAR Lag Order Selection

In the construction of the VAR model, the Lagged differences should be determined first. This is because only by confirming the optimal lag order of the model can exclude the influence of the Granger causality test, variance decomposition and impulse response function on the process of lag order.

TABLE 9
Results of Lagged Differences

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-540.2				65.57	9.859	9.879	9.908
1	-420.9	238.6	4	0	8.059	7.763	7.822	7.90981*
2	-413.6	14.67	4	0.00500	7.586	7.702	7.801	7.947
3	-405.0	17.178*	4	0.00200	6.98012*	7.61848*	7.75788*	7.962
4	-403.5	3.028	4	0.553	7.306	7.664	7.843	8.106
5	-403.0	1.032	4	0.905	7.788	7.727	7.946	8.267
6	-401.6	2.702	4	0.609	8.180	7.775	8.034	8.413
7	-399.3	4.597	4	0.331	8.447	7.806	8.105	8.543
8	-395.6	7.378	4	0.117	8.508	7.812	8.150	8.646

Table 9 shows the results of six evaluation measures. The optimal lag order of LR FPE AIC SBIC HQIC is order 3. On top of that, this paper chooses to build VAR (3) model. Note: * means Optimal lag order of information criterion.

4.4 Cointegration Test

The cointegration test is based on the single integration of the same order, so this paper carries out a cointegration analysis in the following. It should be noted here that the cointegration results only indicate whether there is a long-term equilibrium relationship between variables. After the ADF test shows that RET and RET sequence are in the first-order single integration state, the cointegration test is carried out to determine whether there is a cointegration relationship, which means a long-term equilibrium relationship between them. Our study uses a trace test and the results are shown in Table10.

TABLE 10

Results of Cointegration Test

rank	parms	LL	eigenvalue	trace statistic	5% critical value
0	6	-476.1	.	75.42	15.41
1	9	-440.4	0.460	3.925	3.760
2	10	-438.4	0.0333		

Table 10 shows the cointegration trace test results indicating that there is no long-term equilibrium relationship between sequences.

It can be seen from the trace test that the null hypothesis without a co-integration equation cannot be significantly rejected at the significance level of 5%, and the null hypothesis of at most 0 cointegration equations can be accepted, indicating that there is no long-term equilibrium relationship between sequences.

4.5 Granger Causality Test

Since the SENT and RENT do not have a long-term equilibrium relationship, the Granger Causality Test is conducted to further study their relationship. We then use the Granger causality test to analyse the relationship between the comprehensive index of investor sentiment and the overall return of China's stock market. The specific results are shown in Table 11.

TABLE 11

Granger Causality Test Results

Equation	Excluded	F	df	df r	Prob>F
RET	SENT	0.740	3	108	0.530
RET	ALL	0.740	3	108	0.530
SENT	RET	6.660	3	108	0.0004
SENT	ALL	6.660	3	108	0.0004

Table 11 Granger Causality Test is mainly used to analyse whether there is a statistical causality between economic variables in the model. Results show that RET does not Granger cause SENT.

As can be seen from Table 11, for the null hypothesis of "RET does not Granger cause SENT change", the corresponding p-value is 0.53, which is greater than 0.05, indicating that the null hypothesis cannot be rejected at the significance level of 5%. Therefore, RET does not Granger cause SENT. For the original hypothesis of "SENT does not Granger cause RET change", the corresponding p-value is 0.0004, less than 0.05, indicating that the null hypothesis is significantly rejected at the significance level of 5%. It means that SENT is the Granger cause of RET change.

To sum up, investor sentiment is the Granger cause of stock return, while the stock return is not the Granger cause of investor sentiment.

5. Conclusion

Based on the analysis of the research literature on investor sentiment and the characteristic of China's stock market, this paper selects the proxy indicator of investor sentiment from the transaction data of China's stock market through the analysis of the influencing factors and selection principles of investor sentiment. Due to the correlation among the variables, this paper constructs a comprehensive index to reflect the market sentiment in China by principal component analysis. Our findings suggest that the sentiment factor is a significant factor for both investors and firms to consider when they make investment activities. Finally, the paper investigates the impact of investors on stock return volatility through Impulse Response and Variance Decomposition analysis.

First, this paper summarizes alternative indicators to measure investor sentiment including Closed-end Fund Discount, Turnover Rate, New Investor Accounts, Trading Volume, P/E Ratio, Margin Trading, and Short Sell. Using the principal component regression analysis method, we establish a comprehensive index to reflect investor sentiment. The effectiveness of the comprehensive index and the overall return rate of China's stock market is also tested to suggest that the comprehensive sentiment index of investor sentiment constructed in this paper is a reasonable and effective index.

Second, the Granger causality test suggests that investor sentiment is the Granger cause of the overall return change in China's stock market, and stock return is not the Granger cause of investor sentiment change. Moreover, through the VAR regression analysis between them, the stock returns in the lag period and investor sentiment in the lag period have no significant predictive effect on the current return rate.

Finally, we find that the relationship between investor sentiment and stock return in China does not exist for the long-term situation through Cointegration Test. Furthermore, investor sentiment only explains a small portion of stock return change and shows that investor sentiment has little effect on stock market returns are found by conducting the Impulse and Variance Decomposition analysis.

References

- Bai, M. (2021). Are firm characteristics priced differently between opposite short-sales regimes?. *International Finance*, 24(1), 95-118.
- Bai, M., Li, X. M., & Qin, Y. (2017). Shortability and asset pricing model: Evidence from the Hong Kong stock market. *Journal of Banking & Finance*, 85, 15-29.
- Baker, M., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets*, 7(3), 271-299.
- Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4), 1645-1680.
- Baker, M., Wurgler, J., & Yuan, Y. (2012). Global, local, and contagious investor sentiment. *Journal of Financial Economics*, 104(2), 272-287.
- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, 9(1), 3-18.
- Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307-343.
- Black, F. (1986). Noise. *The Journal of Finance*, 41(3), 528-543.
- Brown, G. W., & Cliff, M. T. (2005). Investor sentiment and asset valuation. *The Journal of Business*, 78(2), 405-440.
- Daniel, K. D., Hirshleifer, D., & Subrahmanyam, A. (2001). Overconfidence, arbitrage, and equilibrium asset pricing. *The Journal of Finance*, 56(3), 921-965.
- Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *The Journal of Finance*, 53(6), 1839-1885.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4), 703-738.
- Drakos, K. (2010). Terrorism activity, investor sentiment, and stock returns. *Review of Financial Economics*, 19(3), 128-135.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- Fisher, K. L., & Statman, M. (2000). Investor sentiment and stock returns. *Financial Analysts Journal*, 56(2), 16-23.
- French, K. R. (1980). Stock returns and the weekend effect. *Journal of Financial Economics*, 8(1), 55-69.
- Gibbons, M. R., & Hess, P. (1981). Day of the week effects and asset returns. *Journal of Business*, 579-596.
- Glushkov, D. (2006). Sentiment beta. Available at SSRN 862444.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417.
- Lee, C. M., Shleifer, A., & Thaler, R. H. (1991). Investor sentiment and the closed-end fund puzzle. *The Journal of Finance*, 46(1), 75-109.

- Li, H., Guo, Y., & Park, S. Y. (2017). Asymmetric Relationship between Investors' Sentiment and Stock Returns: Evidence from a Quantile Non-causality Test. *International Review of Finance*, 17(4), 617-626.
- Ljungqvist, A., Nanda, V., & Singh, R. (2006). Hot markets, investor sentiment, and IPO pricing. *The Journal of Business*, 79(4), 1667-1702.
- Markowitz, H. M. (1952). (1952). Portfolio selection. *Journal of Finance*, 7(1), 77-91.
- Rajnish, M., & Prescott, E. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2), 145-161.
- Schneller, D., Heiden, S., Heiden, M., & Hamid, A. (2018). Home is Where You Know Your Volatility–Local Investor Sentiment and Stock Market Volatility. *German Economic Review*, 19(2), 209-236.
- Shiller, R. J. (2005). Behavioral economics and institutional innovation.
- Shleifer, A. (2000). *Inefficient markets: An introduction to behavioural finance*. OUP Oxford.
- Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, 52(1), 35-55.
- Thaler, R. H. (1999). The end of behavioral finance. *Financial Analysts Journal*, 55(6), 12-17.
- Zweig, M. E. (1973). An investor expectations stock price predictive model using closed-end fund premiums. *The Journal of Finance*, 28(1), 67-78.

Appendix – Other Tests

A.1 Stability test of VAR model

As shown in Figure 2, the unit circle curve of the model and the location map of all the characteristic roots of the VAR model can be seen. All the characteristic roots in the model fall within the unit circle and do not exceed the characteristic roots of the unit circle. This shows that the model is stable and can be used for pulse analysis and variance decomposition.

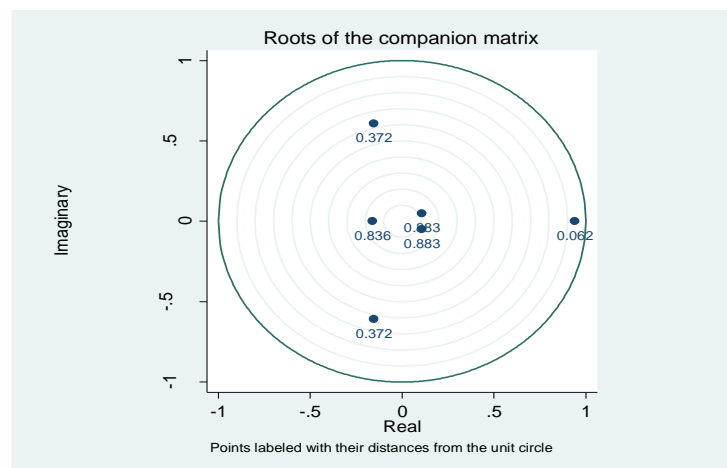


Figure 2. AR diagram of VAR model of investor sentiment

Figure 2 is the unit circle curve of the model showing the location map of all the characteristic roots of VAR model. All the characteristic roots in the model fall within the unit circle and do not exceed the characteristic roots of the unit circle. This shows that the model is stable and can be used for pulse analysis and variance decomposition.

A.2 Impulse Response

The impulse response function reflects the dynamic influence of other variables in the VAR model when one variable is impacted by exogenous shock. According to the dynamic changes of these variables in a period after the impact, the pulse response graph is drawn. This paper further analyses the relationship between stock return and investor sentiment by impulse response analysis. The results are shown in Figure 3 and Figure 4.

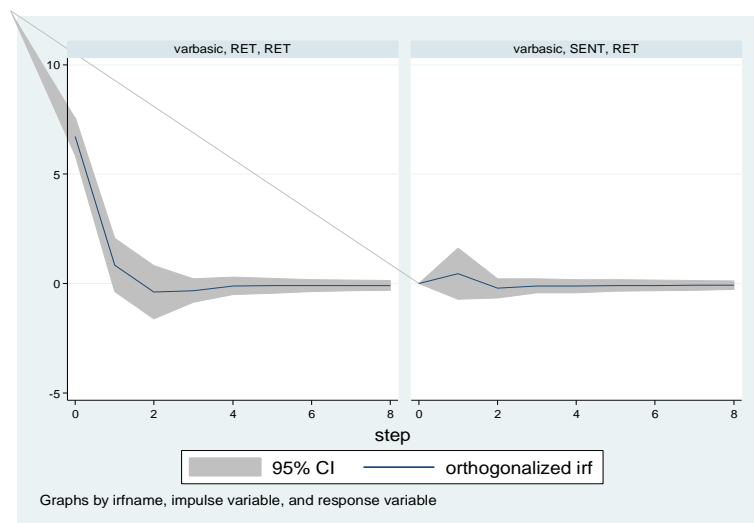


Figure 3. Impulse Response Results of RET

Figure 3 shows the pulse response results for RET. We can find when RET is given a positive impact, RET will rise sharply and then drop rapidly in the first period, indicating that investor sentiment is positively related to the stock return in a short term.

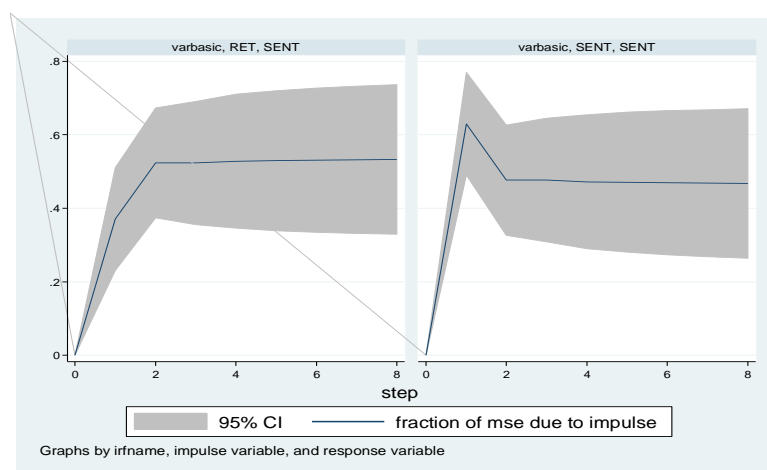


Figure 4. Impulse Response Results of SENT

Figure 4 shows the pulse response to SENT. We can find when a positive impact is given by SENT, the SENT will rise significantly and then decline in the first two periods and then stabilize after the third period, indicating the SENT has a positive impact on itself.

Specifically, figure 3 shows the pulse response results for RET. As can be seen from the figure on the left, when RET is given a positive impact, RET will rise sharply and then drop rapidly in the first period. After that, it will return to about 0 to stabilize. This

indicates that RET has self-correlation and market contagion. It can be seen from the figure on the right that when a positive impact is given to SENT, RET first rises in the first period and then decreases, and finally tends to stabilize after the third period. This indicates that investor sentiment will positively affect the stock return in China in a short term.

Figure 4 shows the pulse response to SENT. As can be seen from the figure on the left, the SENT in the next period will increase significantly if a positive impact is given to RET, and the whole curve above 0 represents a positive impact. The results showed that RET had a significant positive effect on the later SENT. As can be seen from the figure on the right, when a positive impact is given by SENT, the SENT will rise significantly and then decline in the first two periods and then stabilize after the third period.

A.3 Variance Decomposition

Variance decomposition can be used to decompose the variance of a variable into various disturbance items in the VAR model.

TABLE 12
Results of Variance Decomposition

period	RET	SENT
0	0	0
1	1	0
2	0.996	0.00450
3	0.995	0.00543
4	0.994	0.00566
5	0.994	0.00595
6	0.994	0.00611
7	0.994	0.00627
8	0.994	0.00640

Table 12 shows the results of variance decomposition. Variance decomposition test is used to decompose variance of a variable into various disturbance items in VAR model.

It can be seen from Table 12 that with the increase of lag order, RET's ability to explain itself gradually decreases, but overall, it accounts for a large proportion and its variance contribution reaches 99.4% at the 8th stage. The ability of SENT to explain the variance of RET gradually increased, and its variance contribution reached 0.64% at the 8th stage. To conclude, the impact of stock returns on itself gradually decreases, while the impact of investor sentiment increases gradually.