Summer 6-30-2018

The Relationship Between Investor Sentiment and Stock Market Volatility: Based on the VAR Model

Ge Zhang
School of Management Science and Engineering, Shandong University of Finance and Economics, China;School of Information, Capital University of Economics and Business, China

Jishun Wang
School of Management Science and Engineering, Shandong University of Finance and Economics, China

Hao Guo
School of Management Science and Engineering, Shandong University of Finance and Economics, China

Xin Zhang
School of Management Science and Engineering, Shandong University of Finance and Economics, China

The Relationship Between Investor Sentiment and Stock Market Volatility: Based on the VAR Model

Ge Zhang[1,2], *Jishun Wang*[1], *Hao Guo*[1*], *Xin Zhang*[1]
[1]School of Management Science and Engineering, Shandong University of Finance and Economics, China
[2]School of Information, Capital University of Economics and Business, China

Abstract: Using web crawling technology crawls investors' comments of SANY stock(Stock Code: 600031) and Fujian Expressway stock(Stock Code: 600033) from February 11, 2015 to August 16, 2017. Then using semi-supervised machine learning method construct investor sentiment index. Moreover, collecting the daily closing stock price and trading volume data from Qianlong software explore the relationship between investor sentiment and stock market volatility based on VAR model and Granger Test Method. The results show that the rate of return and trading volume have a two-way Granger causality, while negative emotion and the rate of return have a one-way Granger causality. Furthermore, with the impulse response function and variance decomposition, the results show that trading volume has significant effects on rate of return and negative emotions of investors have significant negative effects on rate of return and trading volume.

Keywords: big data, machine learning, investor sentiment, VAR model, stock market volatility

1. INTRODUCTION

In the domestic stock market, there still exist some unreasonable factors, which caused the stock market to rise and fall and bring huge losses to countless investors[1]. Furthermore, the irrational factors of investors increase the uncertainty of investors' decision-making. The traditional standard finance cannot explain the investor's irrational factors influence on the investor well, while the behavior finance explains the irrational behavior of the investor better from the angle of investor's behavior and the psychology of generating such behavior. Investor sentiment as a representative of irrational factors, studying the relationship between investor sentiment and stock market volatility is very important. Domestic and foreign scholars have also made lots of achievements in the field of investor sentiment using traditional financial data[1]-[5].

Recently, with the development and applications of Computer Technology, investors have more opportunities and channels to comment on a particular stock on Internet financial platforms, such as guba.eastmoney.com and guba.sina.com. Those comments are more accurate in reflecting investors' emotional tendencies and gradually becoming an important source of investor sentiment initial data. So how to make better use the comment data and study its role in the stock market has become a focus for many scholars.

This paper uses web crawling technology to collect investors' comments, and quantify the comments to construct investor sentiment index which is an important concept index in finance, then study the interaction between investor sentiment and stock market volatility. The remainder of the paper proceeds as follows. Section 2 introduces the literature review. Section 3 presents the research design. Section 4 analyzes the results. And the last section provides some conclusions and prospect.

2. LITERATURE REVIEW

Many scholars have taken into account the fact that the presence of a large number of emotional-driven investors can lead to price deviations from fundamental value as early as 1936. Research on investor sentiment

* Corresponding author. Email: 422140921@qq.com(Hao Guo)

begins in 1990s, Barberis N etc.[2] aimed at two defects: the empirical model can not effectively response to price information and stock price on the news over reaction, presented a thrifty model based on investor sentiment to study how to form the investors' expectations of future earnings. After 21th century, the research on investor sentiment was gradually increasing, Lee W Y[3] studied the role of investor sentiment in stock market volatility and excess earnings, and concluded that the change of mood was negatively correlated with market volatility and was positively correlated with excess returns. The study of Schmeling M etc.[4] found that emotion has a negative impact on the average returns of stock market in countries. Jiang Yumei etc.[5] discussed the overall and cross-sectional effects of investor sentiment on stock returns. Wang Chun etc.[6] using GARCH-M model to study the influence of investor sentiment on stock market returns and volatility found that there is a positive feedback effect between investor sentiment and stock market returns. Hence, we can know that the impact of investor sentiment on stock returns has been agreed by scholars, but the research conclusions based on different emotional measurement indicators are differently in direction and extent.

Investor sentiment measures can be divided into three parts by foreign scholars: direct, indirect and meta methods[7]. Since meta method is a hybrid version of direct method and indirect method, this paper adopts the viewpoint that there are two methods: direct method and indirect method, which are accepted by most domestic scholars, to measure investor sentiment. The first method based on investors’ survey data about future market trends, e.g. American Association of individual investors Index, CCTV system index, HaoDan Index. Based on above index, Xiong Wei etc.[8] adopt the HaoDan Index analyzing the dynamic relationship between stock trait volatility, stock returns and investor sentiment from the perspective of theory and practice. Zhang Ziqiong etc.[9] using the investor opinion survey data of Sina Finance and Economics stock column calculated the investors’ sentiment, then took the stock of the Shanghai 180 index as the research object to carry out an empirical study. Although emotional data which was gotten though this method needs less to process, and was convenient to use, the cost of obtaining is high and it is hard to process. Currently, the research on investor sentiment is mainly based on the second methods, the proxy index of which includes CEFD, retail selling and buying stocks ratio, CCI[10]. Liu Weiqi etc.[11] select the monthly new accounts of individuals and institutions as individual and institutional investors’ emotional proxy indicators, making a comparative study of the two effects to determine the role of the two emotions in the market.

With the popularization of big data technology, some scholars use crawler program to grab investor’ comments on Internet platform, get investor sentiment index by quantification, and then study the relationship between investor sentiment and stock market. Lai Kaisheng etc.[12] grabbed comments from Sina Weibo, forming Weibo emotion comprehensive index through a series of processes, and discussed the relationship between emotion and stock market. Shi Yong etc.[13] based on the relevant module data of excellent mining financial platform, the investor attention degree, snowball investor attention degree, news attention degree and news sentiment index were constructed respectively, and used correlation and VAR model to explore the relationship between these indicators and the CSI 300 Index.

In summary, the research hot spot of investor sentiment has not been reduced, and the measurement method has been developed more scientific with the development of technologies. With the literature review, the result shows that using big data technology to build investor sentiment is still in the initial stage. The papers are not common by using big data machine learning to quantify comments as investor sentiment indicators to study related issues. Therefore, this study intends to use the big data machine learning method to build investor sentiment, regarding www.eastmony.com as data sources, to study the relationship between investor sentiment and stock market volatility based VAR model.

3. RESEARCH DESIGN

3.1 Model introduction

Through verification of the initial data found that the VAR model can explain the relationship between each variable and its own lag better. In VAR model, each endogenous variable in the system is regarded as a function of the lag value of all the endogenous variables in the system to construct the model. At the same time, it has less requirements for the characteristics of variables, and could better describe the dynamic behavior in the fields of economy, finance and so on, and it has a good prediction effect. The model will be established in this paper as follows:

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + \sum_{j=0}^{r} B_j X_{t-j} + \varepsilon_t \, , t = 1, 2, \cdots, n \tag{1}$$

Where Y_t is a k dimensional endogenous variable, Y_{t-1} （$i = 1, 2, \cdots, p$） is a lag endogenous variable vector. X_{t-i} （$i = 0, 1, \cdots, r$） is the lag order of a d dimensional exogenous variable or a lag exogenous variable vector, p , r , respectively, is the lag order of the endogenous variable and the exogenous variable. A_i is $k \times k$ dimension matrix, B_j is $k \times d$ dimension matrix, ε_t is a vector of the k dimensional perturbation term.

3.2 Indicator description

Yao Jun etc.[14] pointed out that the volatility of the stock market could be measured by the daily returns of high frequency data. Li Dan etc.[15] drawn an conclusion that the trading volume was positively related to stock market volatility based on the theory of mixed distribution hypothesis. Therefore, this study selects the rate of returns(R), trading volume(V) as proxy indicators for market volatility, and selects the positive investor sentiment (PE) and the negative emotion (PN) as proxy indicators for investor sentiment. The formula for calculating R we adopted as follows:

$$r_t = \ln\left(close_t\right) - \ln\left(close_{t\text{-}1}\right) \tag{2}$$

where, $close_t$ represent the closing price of current trading day.

As for investor sentiment, using big data platform get the index of PE and NE. The processing flow of the big data platform is shown as follows:

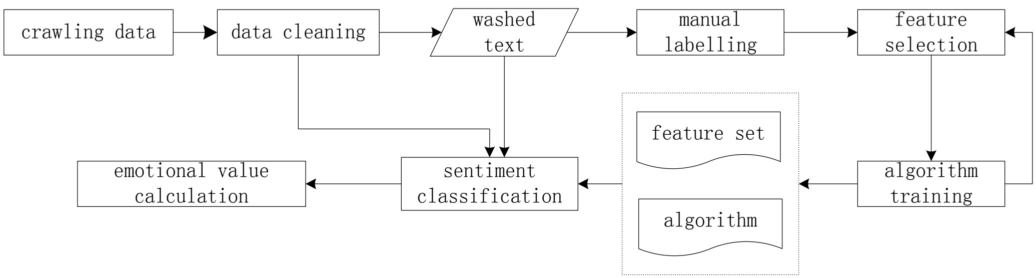

Figure 1. The process and function of the investor sentiment big data platform

3.3 Data source and processing

The data of V and closing price of stock market come from Qianlong software which relying on its products be a pioneer in domestic securities industry. Qianlong software has a long history and contains a variety of time sharing data and technical data.

Dongfang Wealth has the largest access and is the most active website in China's financial and economic securities portal, according to iResearch. Therefore this article takes www.eastmoney.com as the source of initial emotional data. In order to ensure the representative, reliable of data , two stocks (including 600031, 600033) are randomly selected. Using the Python language to write crawl program based on the Pycharm environment crawled comment contents, comment time, amount to be read, amount of comment, commentator's message of two stocks from February 11, 2015 to August 16, 2017 , finally crawled 90342 comments. The process of crawling comments is divided into the following steps :

Step1: data cleaning. In this step, the main task is cleaning the third party institutions' comments and some other messages. After data cleaning, the remaining comments have 87693 pieces of valid.

Step2: sentiment classification. If there is no emotion classifier, an emotion classifier needs to be formed by a series of processes: (1)Inviting 5 investors with many years of stock investment experience manually mark 10,296 valid comments on 600033 stocks. Then, (2)electing of characteristic words has great influence on the accuracy rate of the machine learning algorithm, so it's particular important to choose the right feature words. Next, (3)The training set and test set are selected according to the tagged text and its features. (4)Finally, the emotion classifier is formed according to algorithm and the corresponding feature set, then the 600031 comments are classified.

Step3: emotional value calculation. The emotional index is calculated by using the ratio of the number of comments on one day's positive evaluation (negative evaluation) and the total number of comments on the day, and the emotional index is calculated to get the daily emotional index. Computational formula are as follows:

$$posindex_t = \frac{poscount_t}{total_t} \tag{3}$$

$$negindex_t = \frac{negcount_t}{total_t} \tag{4}$$

Through above methods, the data of each proxy index is obtained, but in order to make every index data with lateral comparability and feasibility, the index value needs to be standardized by Range Standardization Method. At the same time, considering the practical significance of the positive and negative value of the index, the calculation method of the positive index is as follows[16]:

$$x_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \tag{5}$$

When the index is negative, the method of calculation is:

$$x_{ij} = \frac{\max(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})} \tag{6}$$

4. EMPIRICAL ANALYSIS

4.1 Unit root test

After processing of each indicator, the trend of the four indicators are shown in figure 2:

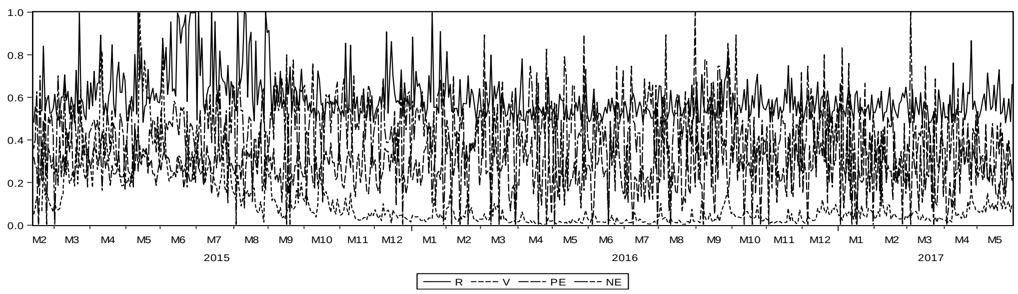

Figure 2. Trend chart of four indicator data

The reliability of the established VAR model depends on the stability of the variables. Stationary time series can be directly constructed non-constraint VAR model. Conversely, non-stationary time series need to check by cointegration test whether it exists a long-term equilibrium relationship between variables. If there is a long-term equilibrium relationship, the vector error correction model could be constructed, otherwise, it is necessary to make the difference of each variable into a stable variable. For further analyze the model, Eviews8 software is used to test the stability of each variable by unit root test. The result as table 1 shows:

Table 1. The result of unit root test

variable	type of models	DF-GLS	ADF	result
R	containing constant term, excluding trend	-2.345**	-6.8236***	stable
V	containing constant term, excluding trend	-3.5491***	-3.805***	stable
PE	containing constant term, excluding trend	-21.8361***	-22.0543***	stable
NE	containing constant term, excluding trend	-6.5702***	-7.7514***	stable

***represents a significant level of confidence at 1%, **represents a significant level of confidence at 5%.

The results of unit root test show that the test statistic values in above four time series are all less than the critical values under the 1% test level, so the four sequences do not contain unit roots, that is, all variables are I(0) stationary variables.

4.2 Construction of VAR model

In this study, four variables are used as endogenous variables to establish VAR model. Through "Lag Structure" function in " Lag Length Criteria", select lag number 7 to carry on analysis, then get table 2.

Table 2. Optimal delay order

Lag	LogL	LR	FPE	AIC	SC	HQ
0	1404.545	NA	1.04e-07	-4.723591	-4.694012	-4.712071
1	1840.797	865.1471	2.53e-08	-6.140967	-5.993069*	-6.083363
2	1890.219	97.34414	2.26e-08	-6.253689	-5.987472	-6.150002
3	1932.550	82.80667	2.07e-08	-6.342497	-5.957960	-6.192726*
4	1952.167	38.10910	2.04e-08	-6.354695	-5.851840	-6.158841
5	1989.987	72.96080*	1.90e-08*	-6.428286*	-5.807112	-6.186349
6	1999.969	19.12167	1.94e-08	-6.407988	-5.668495	-6.119967
7	2009.692	18.49655	1.98e-08	-6.386821	-5.529009	-6.052717

Clearly, the optimal delay order is 5 according to five criteria for determining the optimal delay order. Then, VAR(5) model is established. In order to prove whether VAR(5) model is stable,we make an AR root graph. From the result of AR root graph, as figure 3 indicates, all points are in the unit circle, which is to say the model is stable.

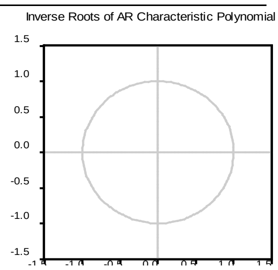

Figure 3. Stability text of VAR model

4.3 Granger causality test

After the VAR model passes the stability test, it needs to test the result of the model estimate using granger method. The final result is shown in Table 3.

Table 3. Granger causality test

Dependent variable: R　　Dependent variable: V

Exduded	Chi-sq	df	Prob.	Exduded	Chi-sq	df	Prob.
V	11.2535	5	0.0466	R	13.0327	5	0.0231
PE	3.3291	5	0.6494	PE	0.1514	5	0.9996
NE	10.6068	5	0.0498	NE	2.58051	5	0.7643
All	27.0212	15	0.0286	All	17.1353	15	0.3108

From the data in Table 3, under the test level of 5%, V and R are mutual Granger reasons. This relationship confirms the phenomenon of stock market. No matter the rise and fall of trading volume and stock price are the actual reflection of supply demand and mapping of investors' psychological behavior. NE is the Grainger reason of R. The reason is that when the investor's mood is low, it would affect the act of selling and buying stocks, and then affect the supply and demand relationship of the stock market, finally affect R. Endogenous variable R is jointly significant relative to the V, PE and NE lag terms, which indicates the lagged variables of V, PE, NE have strong influence on the R. Similarly, the endogenous variable PE is significantly associated with the lagged variables of R, V and NE.

4.4 Pulse response

In order to understand the dynamic characteristics of VAR model, the impulse response function can be used to analyze it. Impulse response function is used to measure the impact of a standard deviation impulse on the current and future values of all endogenous variables in the VAR model. That is to consider how the effect of the residual term spreads to other variables. The real line, in the graph 4(unit: day, lag interval set 30 periods), represents the impulse response function, and the dashed line represents the deviation zone of positive and negative double standard deviation.

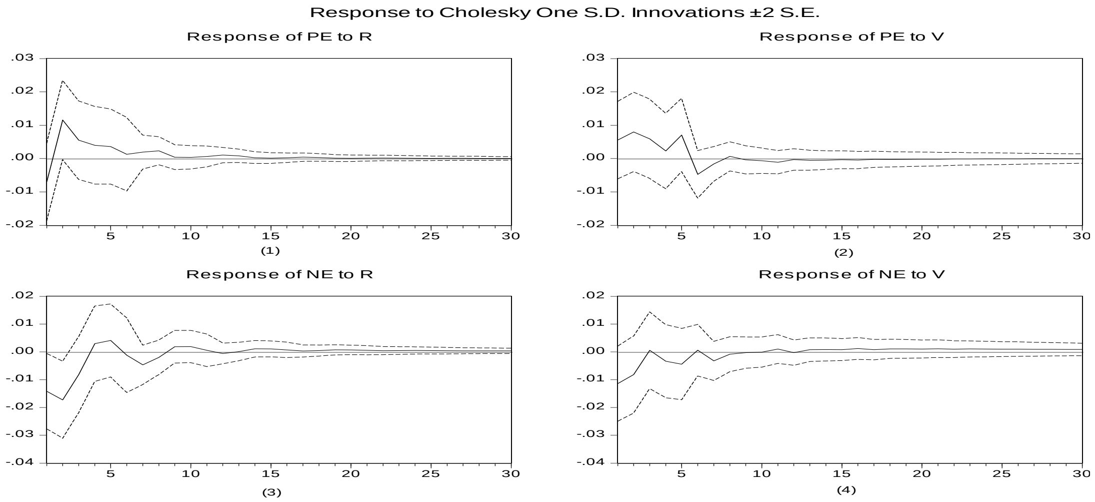

Figure 4. Pulse response of R, V to PE, NE

Figure 4(1) displays that PE will increase in the short term and then return to quiet gradually with the rate of return increasing. Figure 4(3) shows that when R is high, NE decreases in the first two period, but then increases again. It suggests the reason for NE is more complex, while the increase of R is the main reason for PE, which is more clear than the reason for NE. At the same time we find that NE has a greater impact on R and lasts longer.

Figure 4(2) the response of PE to a standard deviation impacting on V is also timely, the value of the response in the first period is about 0.007, and then gradually decreases with the period increasing. From figure 4(4), we can find that the NE in response to a standard deviation shock of V in the first period is -0.11. In a few fluctuations period, it tends to be around 0. The above two cases indicate that the short-term increasement of V will cause emotional differentiation, both positive and negative emotions will increase. Then after going through a large fluctuation, PE return calm.

4.5 Variance decomposition

The impulse function can only analyze the response of some variables to the disturbance of one variable, but can not analyze the importance of exogenous variables to endogenous variables. Therefore, variance decomposition is needed to analyze the contribution of each structural shock to the change of endogenous variables (measured by variance). Figure 6 shows the results of variance decomposition of R and V in the 30 period of delay. The transverse axis represents the number of lag periods, and the longitudinal axis represents the index contribution.

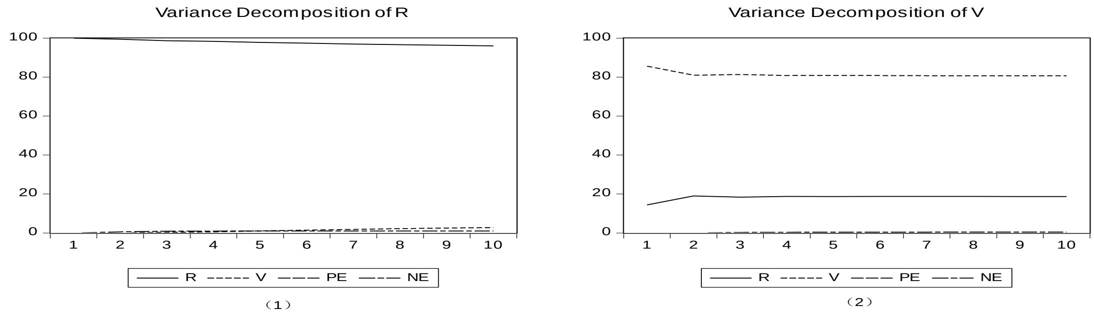

Figure 5. Variance decomposition of R, V index

From Figure 5 (1) we found the contribution of return to itself is 100%, then it decreases gradually in second period while the contribution of trading volume, positive emotion and negative emotion increasegradually. After the fifth period,V, PE and NE's contribution to return arrived 2%, 1%, 3%,respectively. This suggests that the contribution of V, NE to R is significant. From Figure 5 (2), R is highly contribute to V in the first five period at around 23%, PE and NE contribution to the V is insignificant. This phenomenon is consistent with the China stock market which do not allow investors to sell short. When positive emotions are high, on the one hand, there will be a phenomenon of reluctant to sell, on the other hand, the risk will increase, and investors who are willing to buy stocks at a high level will also be reduced causing positive emotions to have little contribution to volume.

5. CONCLUSIONS

Through the method of machine learning construct the investor sentiment index, we studied the mutual influence of investor sentiment and stock market volatility. The results show that there is a two-way Granger causality between trading volume and returns and the negative emotion and the rate of return have one-way Granger causality. Through the impulse response function we find the trading volume to the rate of return is positive, while the response of rate of return to volume is negative.

The main contribution of this paper is that using machine learning method to construct investor sentiment index, under the background of big data, combined with VAR model of econometrics to study the relationship between investor sentiment and stock market volatility. The results confirm the conclusions of previous scholars' research, indicating that the method of constructing investor sentiment index is feasible and effective. However, there are several points in this paper that need to be continued in the future research.

1. In this paper, big data technique is used to analyze stock comments, quantifying the text into emotional values. In the process of research, emotional values may deviate from actual values due to the subjectivity of manual annotation. There is a certain impact on the results of the research. In the future research, the comments of investors will be dealt with more persuasively. For example, more senior stock investors and stock researchers will be invited to manually annotate comments to form a more effective training set.

2. From the results of this paper, we can know that the positive and negative emotions have a significant impact on the stock returns and trading volume in different ways. The next step is to expand the sample size, adopt the data of different stock market and more stocks.

3.Though the impulse response function find the change of positive emotions and negative emotions are not the relationship of one is rising and the other is falling, sometimes it can also be strengthened or weakened at the same time. In the future research, sentiment of investors will be subdivided more detailedly to explore the stock market investors emotional impact on stock market volatility.

ACKNOWLEDGEMENT

This research was supported by the Natural Science Foundation of Shandong Province under Grant ZR2012GQ006 and the the National Social Science Foundation of China under Grant 17BGL198 and 16AJY003.

REFERENCES

[1] Lai Kaisheng, Chen Hao, Le Guoan. (2014). Can mood predict the stock market? Advances in Psychological Science. 22(11):170-1781.

[2] Barberis N, Shleifer A, Vishny R W. (1997). A Model of Investor Sentiment. Journal of Financial Economics. 49(3):307-343.

[3] Lee W Y, Jiang C X, Infro D C. (2002). Stock Market Volatility, Excess Returns, and the Role of Investor Sentiment. Journal of Banking & Finance. 26(12):2277-2299.

[4] Schmeling M. (2009). Investor sentiment and stock returns: Some international evidence. Journal of Empirical Finance. 16(3): 394-408.

[5] Jiang Yumei, Wang Mingzhao. (2010). Investor sentiment and stock return : an empirical study of overall effect and cross-sectional effect. Nankai Business Review. 13(3):150-160(in Chinese).

[6] Wang Chun. (2014-9). The influence of investor sentiment on the returns and volatility of the stock market-An Empirical Study Based on the net inflow of open stock fund money.Chinese Journal of Management Science. 22(9):49-56(in Chinese).

[7] Kim T, Ha A. (2010). Investor Sentiment and Market Anomalies. 23rd Australasian Finance and Banking Conference 2010 Paper.

[8] Xiong Wei, Chen Langnan. (2015-9). Stock Price Volatility , Stock Returns and Investor Sentiment. Journal of Management Science. 28(5):106-115(in Chinese).

[9] Zhang Ziqiong, Qian Guoming, Li Yijun. (2013-7). An empirical study on the interaction between investor sentiment and Chinese stock market based on online voting. Soft Science. 27(7): 123-127(in Chinese).

[10] Zhang Qiang, Yang Shue, Yang Hong. (2007-7). An empirical study on investor sentiment and stock returns in Chinese stock market. Systems Engineering. 25(7): 13-17(in Chinese).

[11] Liu Weiqi, Liu Xinxin. Individual and institutional investor sentiment and stock returns: a study based on Shanghai Stock Market. Journal of Management Sciences in China. 17(3):70-87(in Chinese).

[12] Lai Kaisheng, Chen Hao. (2014-5). Weibo emotion and Chinese stock market : Based on Cointegration Analysis. Journal of Systems Science and Mathematical Sciences. 34(5): 565-575(in Chinese).

[13] Shi Yong, Tang Jing, Guo Kun. (2017-7). The concern of social media investors and investor sentiment influence on the Chinese stock market. Journal of Central University of Finance & Economics. 45-53(in Chinese).

[14] Yao Jun, Yin Kairui, Li Qi. (2017-11). The impact of Macro fundamentals and institutional investors on the volatility of stock market : a review and prospect. Modernization of Management. (6): 118-121(in Chinese).

[15] Wen Xi, Wang Guoshun. (2015-3). Research on the dynamic relationship between liquidity and returns of fuel oil futures market. Systems Engineering. 33(3): 38-44(in Chinese).