

Construction of Investor Sentiment Index in the Chinese Stock Market

Yuxi Yang ^{*}, Takashi Hasuike [†],

Abstract

This paper focuses on improving the adaptability of the investor sentiment index introduced by Baker and Wurgler in the Chinese stock market. Considering not all the original proxies for sentiment are suitable for Chinese stock market, a new combination of proxies to form the investor sentiment index is proposed. Based on this investor sentiment index, the relationship between investor sentiment and stock price index in Chinese stock market is found.

Keywords: Investor sentiment, Chinese stock market, behavioral finance.

1 Introduction

According to the efficient market hypothesis and economic man hypothesis, all the investors are expected to act perfectly rationally to the newly disclosed information. Therefore, securities prices are always reasonably estimated as new net present values of cash flows. However, in reality, abnormal skyrocketing and collapse in stock price have been occurring all the time. Such a phenomenon is thought to be difficult to explain in the classical finance theory. On the other hand, in the view of behavioral finance which is developed rapidly in recent years, investor sentiment is considered to be one of the factors affecting stock price. As for the application, the investor sentiment could be used as a factor in the multi-factor model to predict stock returns. For these reason, researches on investor sentiment are becoming more and more important.

As the largest emerging country, the scale of Chinese stock market is second only to the United States becoming the second in the world. Characteristics of the Chinese stock market include a high percentage of individual investors who are susceptible to investor sentiment, and many speculative transactions. Therefore, the influence of investor sentiment on the Chinese stock market is thought to be larger than other mature stock markets such as the United States.

As a representative study on investor sentiment, Baker and Wurgler [1] can be mentioned. Considering it was hard to measure investor sentiment quantitatively in conventional researches, Baker and Wurgler [1] used the principal component analysis method to compress several proxies for investor sentiment and synthesized the investor sentiment index (BW method). After that study, the BW method became the most popular measurement to quantify investor sentiment.

^{*} Waseda University, Tokyo, Japan

[†] Waseda University, Tokyo, Japan

However, not all the proxies used in BW method are suitable for Chinese stock market. For example, in the Chinese stock market, since there is an examination system for IPO, proxies about IPO are not qualified to be applied. Therefore, we propose a new combination of proxies to form the investor sentiment index which is more suitable for Chinese stock market. Based on this investor sentiment index, we also analyze the influence of investor sentiment on the Chinese stock market.

2 Previous Studies

In the previous researches, there are mainly two methods to measure investor sentiment. The first method is to ask investors directly about what they think about the future market by questionnaire. Count up the numbers of people in the buying position and the selling position, and the ratio of the two numbers is the measurement to quantify investor sentiment. A representative study on this method is Fisher and Statman [2]. The second method is to use a specific stock index as a substitute indicator for measuring investor sentiment. For example, Lee et al. [3] proposed that the closed-end fund discount could be used to measure investor sentiment.

However, either of the methods is limited by the amount of information, and it is impossible to measure investor sentiment comprehensively. Baker and Wurgler [1] uses the principal component analysis method to extract common element of investor sentiment from several stock indices, and synthesize the investor sentiment index. The investor sentiment index generated by the BW method can comprehensively evaluate the investor sentiment and is superior to conventional methods.

2.1 BW Method

The six proxies for investor sentiment used in BW method are listed as follows.

- The closed-end fund discount (CEFD)
- NYSE share turnover (TURN)
- The number of IPOs (NIPO)
- The average first-day returns (RIPO)
- The share of equity issue in total equity and debt issues (S)
- The dividend premium (P)

Baker and Wurgler [1] also claimed that there is a lead-lag effect between various investor sentiment indicators. The effect of the lead-lag is that different investor sentiment indicators have a time lag in the reflection speed for investor sentiment.

For example, it is believed that such a lead-lag effect exists between the number of IPOs and the average first-day returns. The reason for this is that the investors' willingness to invest will also rise as average first-day returns increase, resulting in more IPOs. Therefore, the average first-day returns reflect investor sentiments before the number of IPOs.

Baker and Wurgler [1] used two-stage principal component analysis in order to clarify this lead-lag effect. In this study, we took the same steps to clarify the lead-lag effect.

2.2 Problem of Previous Studies

For some reasons, not all the proxies used in BW method is suitable for Chinese stock market. For example, in Chinese stock market, IPOs are severely limited by the China Securities Regulatory Commission. Therefore, it is considered that the indices related to IPO, such as the number of IPOs and the average first-day returns of IPO stocks, are not supposed to reflect investor sentiment correctly.

3 The Selection of Proxies for Investor Sentiment

To improve the adaptability of the investor sentiment index generated by the BW method, we propose a new combination of proxies to form the investor sentiment index. The followings are the proxies.

- The closed-end fund discount (CEFD)
- The number of new account opening (OPEN)
- Market turnover (TURN)
- Margin debt (MD).

The reasons for using them are shown in the following sections. Many previous researches on measuring investor sentiment uses monthly data, but in this paper, in order to measure the investor sentiment more precisely, we used the weekly data. The data period is from January 7, 2011 to October 28, 2016. (The source of all data is from WIND database). Also, in order to eliminate the influence of data units, all the investor sentiment proxies are standardized.

3.1 Closed-end Fund Discount

The closed-end fund discount is expressed as the difference between the net asset price (NAV) of the closed-end fund and the purchase price (P). Lee et al. [3] said that CEFD can be used as an indicator to measure investor sentiment. Neal and Wheatley [4] and Fisher and Statman [2] also indicated that CEFD is effective as an index to measure investor sentiment. Baker and Wurgler [1] used the CEFD as a variable to represent investor sentiment. Normally, it is said that CEFD have a negative correlation with investor sentiment.

This study used a weighted average taken from six close-end funds to represent the CEFD of the whole market. The six closed-end funds are the YINFENG Fund, the KERUI Fund, the HONGYANG Fund, the JIUJIA Fund, the FENGHE Fund and the TONGQIAN Fund.

3.2 Number of New Account Opening

The number of new account opening can intuitively reflect changes of investor sentiment by measuring the degree of participation in investment activities. Hu and Chi [5] showed that it is effective as an indicator to measure investor sentiment. Normally, it is said to have a positive

correlation with investor sentiment.

3.3 Market Turnover

The market turnover is an indicator showing the activity of the stock market. It is calculated by dividing the trading volume of each period by market capitalization of the same period. Baker and Wurgler [1] and Hu and Chi [5] used the TURN as a representative variable of investor sentiment, when synthesizing the investor sentiment indices. Normally, it is said to have a positive correlation with investor sentiment.

3.4 Margin Debt

The margin debt indicates the outstanding loans from financial institutions. On March 31, 2010, the Chinese stock market launched the margin transaction called "Rongzirconquan". Because of the short history of margin transaction, research has not been promoted much yet. But its influence can be considered to be huge. Because it is believed that the Chinese stock market crash of mid-2015 was caused by credit transactions.

4 The Construction of Investor Sentiment Index

According to the lead-lag effect between the investor sentiment indices, it is necessary to use two-stage principal component analysis to clarify the lead-lag relationship between investor sentiment proxies.

In the first stage of principal component analysis, we used the four investor sentiment proxies and the data before one term of each the four proxies. Based on the eight variables, a provisional investor sentiment index (tSENT) can be obtained.

Next, figure out the Pearson correlation coefficients between the eight variables with the provisional investor sentiment index. Select the variables with a higher correlation coefficient and use them as the variables to obtain the formal investor sentiment index (SENT).

4.1 The Construction of the Provisional Investor Sentiment Index

4.1.1 KMO and Bartlett's Test

Before conducting the first-stage principal component analysis, in order to verify the compatibility of usage data with principal component analysis, it is necessary to conduct KMO and Bartlett's test. The results are shown in Table 1. According to the Table 1, the result of the KMO is 0.741 and the significant probability of the Bartlett sphericity test is 0.000. Therefore, it is considered that there is a high correlation in the different variables and the variables are suitable for principal component analysis

Table 1: KMO and Bartlett's Test (tSENT)

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.741
	Approx. Chi-Square	4772.251
	Degree of Freedom	28
	Significant Probability	0.000

4.1.2 Principal component analysis

Perform principal component analysis on the eight variables. The results are shown in Table 2 and Table 3. According to Table 2, considering the percentage of the contribution ratio, the first principal component and the second principal component can be used to represent the provisional investor sentiment index. According to Table 3, considering the coefficient of the variables, the first principal component is qualified to represent the provisional investor sentiment index.

Table 2: Eigenvalue and Contribution Ratio (tSENT)

Principal Component	Eigenvalue	Contribution Ratio	Cumulative Contribution Ratio
1	5.366	67.078	67.078
2	1.926	24.070	91.148
3	0.472	5.905	97.053
4	0.091	1.137	98.190
5	0.076	0.949	99.139
6	0.044	0.551	99.690
7	0.023	0.281	99.971
8	0.002	0.029	100.000

Table 3: Coefficient of Variables (tSENT)

Variables	Principal Component	
	1	2
CEFD	-0.043	0.499
OPEN	0.166	-0.022
TURN	0.178	0.097
MD	0.178	0.034
L.CEFD	-0.045	0.502
L.OPEN	0.167	-0.006
L.TURN	0.177	0.102
L.MD	0.176	0.036

4.2 The Selection of Investor Sentiment Proxies Variables

According to Table 3, the provisional investor sentiment index (tSENT) is defined as follows.

$$\begin{aligned}
 tSENT_t = & -0.043CEFD + 0.166OPEN \\
 & + 0.178TURN + 0.178MD \\
 & - 0.045L.CEFD + 0.167L.OPEN \\
 & + 0.177L.TURN + 0.176L.MD
 \end{aligned} \tag{1}$$

Select the variables with higher absolute values of correlation coefficients and use them as the variables of formal investor sentiment index. As a result, L.CEFD, L.OPEN, TURN, MD are selected.

4.3 The Selection of Investor Sentiment Proxies Variables

4.3.1 KMO and Bartlett's Test

Before conducting the second-stage principal component analysis, in order to verify the compatibility of usage data with principal component analysis, it is necessary to conduct KMO and Bartlett's test. The results are shown in Table 4. According to the Table 4, the result of the KMO is 0.622 and the significant probability of the Bartlett sphericity test is 0.000. Therefore, it is considered that there is a high correlation in the different variables and the variables are suitable for principal component analysis.

Table 4: KMO and Bartlett's Test (SENT)

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.622
	Approx. Chi-Square	981.826
	Degree of Freedom	6
	Significant Probability	0.000

4.3.2 Principal component analysis

Perform principal component analysis on the four variables. The results are shown in Table 5 and Table 6. According to Table 5, considering the percentage of the contribution ratio, the first principal component and the second principal component can be used to represent the formal investor sentiment index. According to Table 6, considering the coefficient of the variables, the first principal component is qualified to represent the formal investor sentiment index.

According to Table 6, the formal investor sentiment index (SENT) is defined as follows.

$$SENT_t = -0.043 L.CEFD + 0.167 L.OPEN + 0.178 TURN + 0.178 MD \quad (2)$$

Table 5: Eigenvalue and Contribution Ratio (SENT)

Principal Component	Eigenvalue	Contribution Ratio	Cumulative Contribution Ratio
1	2.718	67.939	67.939
2	0.981	24.533	92.472
3	0.248	6.207	98.679
4	0.0528	1.321	100

Table 6: Coefficient of Variables (SENT)

Variables	Principal Component	
	1	2
L.CEFD	-0.088	0.997
L.OPEN	0.334	-0.018
TURN	0.351	0.193
MD	0.350	0.069

5 Granger Causality

5.1 Unit Root Test

To find out the relationship between the investor sentiment and stock market index, it is necessary to clarify the granger causality between them.

But first of all, in order to avoid the spurious regression, before conducting the granger causality test, it is necessary to conduct the unit root tests.

According to Table 7, it is considered that unit roots are existing in this time series at the significance level of 5%. Therefore, in order to get a more accurate result, it is need to conduct a unit root test for the first-order differentiation of SENT.

According to Table 8 it is considered that there is no unit root exists in this time series even at the significance level of 1%. Next, the results of unit root test for Shanghai Composite Index (INDEX) are shown in Table 9.

According to Table 9, it is considered that unit roots exist in this time series at the significance level of 10%. Therefore, in order to get a more accurate result, it is need to conduct a unit root test for the first-order differentiation of INDEX.

According to Table 10, it is considered that there is no unit root exists in this time series even at the significance level of 1%. As there is no unit root exist in the first-order Differentiation of INDEX and SENT, it is possible to conduct the granger causality test between SENT and INDEX.

As for the average returns (RETURN), it has the same results with the first-order differentiation of INDEX. Therefore, it is also considered that there is no unit root exist in RETURN at the significance level of 1%.

Table 7: Unit Root Test (SENT)

	T-Statistic	Probability
Augmented Dickey-Fuller Test Statistic	-1.79957	0.0685
1% level	-2.57291	
5% level	-1.94191	
10% level	-1.61597	

Table 8: Unit Root Test
(First-order Differentiation of SENT)

	T-Statistic	Probability
Augmented Dickey-Fuller Test Statistic	-10.177	0.000
1% level	-2.57291	
5% level	-1.94191	
10% level	-1.61597	

Table 9: Unit Root Test (INDEX)

	T-Statistic	Probability
Augmented Dickey-Fuller Test Statistic	-1.44297	0.139
1% level	-2.57282	
5% level	-1.9419	
10% level	-1.61598	

Table 10: Unit Root Test
(First-order Differentiation of INDEX)

	T-Statistic	Probability
Augmented Dickey-Fuller Test Statistic	-15.1342	0.000
1% level	-2.57291	
5% level	-1.94191	
10% level	-1.61597	

5.2 Granger Causality Test

To obtain a more reliable result of granger causality test, before conducting the granger causality test, it is necessary to find out the best lag length by referring to the Bayesian Information Criterion. As what are shown in Table 11 and Table 12, based on the Bayesian Information Criterion, the best lag length for granger causality test is 2.

Table 11: BIC (SENT and INDEX)

Lag	BIC
0	4.048363
1	-1.447244
2	-1.584917*
3	-1.533525
4	-1.515058
5	-1.448414
6	-1.386488
7	-1.438283
8	-1.382668

Table 12: BIC (SENT and RETURN)

Lag	BIC
0	5.744361
1	2.024341
2	2.020396*
3	2.07884
4	2.103218
5	2.170251
6	2.233627
7	2.22473
8	2.272576

The results of granger causality test for SENT and INDEX are shown in Table 13. According to Table 13, at the level of lag length 2, it is considered that SENT granger causes INDEX and INDEX also granger causes SENT at the significance level of 1%.

The results of granger causality test for SENT and RETURN are shown in Table 14.

According to Table 14, at the level of lag length 2, it is considered that RETURN granger causes SENT at a significance level of 1%. Also, SENT granger causes RETURN at the significance level of 5%.

Table 13: Granger Causality Test (SENT and INDEX)

Null Hypothesis:	F-Statistic	Probability
SENT does not Granger Cause INDEX	11.8997	1.00E-05
INDEX does not Granger Cause SENT	7.44745	0.0007

Table 14: Granger Causality Test (SENT and RETURN)

Null Hypothesis:	F-Statistic	Probability
SENT does not Granger Cause RETURN	3.03363	0.0497
RETURN does not Granger Cause SENT	9.34578	0.0001

6 Empirical Results and Consideration

6.1 Empirical Results

6.1.1 Linear Model

First, to test the effectiveness of the investor sentiment index under the linear model, we used the LS method and the model is defined as follows.

$$INDEX_t = c_t + \alpha_0 SENT_t + \varepsilon_t \quad (3)$$

$$RETURN_t = c_t + \alpha_0 FOSENT_t + \varepsilon_t \quad (4)$$

Where c is a constant, ε is a stochastic error, and α is the Pearson correlation coefficient. According to Table 15, the Pearson correlation coefficient between SENT and INDEX is 0.893. Therefore, it is considered that SENT can fully reflect the fluctuations of INDEX.

According to Table 16, the Pearson correlation coefficient between the first-order Differentiation of SENT and RETURN is 0.395. Therefore, it is considered that there is no strong correlation between the first-order Differentiation of SENT and RETURN.

Table 15: Pearson Correlation Coefficient (SENT and INDEX)

Pearson correlation coefficient	SENT	INDEX
SENT	1	0.893
INDEX	0.893	1

Table 16: Pearson Correlation Coefficient (First-order Differentiation of SENT and RETURN)

Pearson correlation coefficient	first-order Differentiation of SENT	RETURN
first-order Differentiation of SENT	1	0.395
RETURN	0.395	1

6.1.2 Nonlinear Model

Second, to have more extensive applicability, it is necessary to test the effectiveness of the investor sentiment index under the nonlinear model. Using the GARCH(1,1), GARCH(2,1) and GARCH(1,2) which are the most common GARCH models to test the suitability of investor sentiment.

The results of GARCH(1,1), GARCH(2,1) and GARCH(1,2) model are shown in Table 17, Table 18 and Table 19. According to Table 17, Table 18 and Table 19, under the GARCH(1,1) model, the Significant Probability of variables has the best performance.

Table 17: GARCH(1,1)

Variable	Coefficient	Std. Error	z-Statistic	Probability
SENT	0.800118	0.018022	44.395950	0.000000
	Variance Equation			
C	0.005059	0.002055	2.461522	0.013800
RESID(-1)^2	0.486107	0.133988	3.627992	0.000300
GARCH(-1)	0.492291	0.062986	7.815895	0.000000

Table 18: GARCH(2,1)

Variable	Coefficient	Std. Error	z-Statistic	Probability
SENT	0.818401	0.017107	47.839480	0.000000
	Variance Equation			
C	0.001911	0.001981	0.964533	0.334800
RESID(-1)^2	0.706883	0.213242	3.314934	0.000900
RESID(-2)^2	-0.508260	0.266169	-1.909540	0.056200
GARCH(-1)	0.789271	0.166197	4.749002	0.000000

Table 19: GARCH(1,2)

Variable	Coefficient	Std. Error	z-Statistic	Probability
SENT	0.845499	0.017308	48.850840	0.000000
	Variance Equation			
C	0.007381	0.003359	2.197589	0.028
RESID(-1)^2	0.56265	0.176536	3.187173	0.0014
GARCH(-1)	0.151152	0.326281	0.463258	0.6432
GARCH(-2)	0.235449	0.254792	0.924083	0.3554

Based on the result of the GARCH(1,1) model, it is necessary to consider the suitability of investor sentiment index to T-GARCG(1,1) model and the E-GARCG(1,1).

The results of T-GARCG(1,1) model and the E-GARCG(1,1) are shown in Table 20 and Table 21. According to Table 20 and Table 21, not all the Variables are suitable under the T-GARCG(1,1) model and the E-GARCG(1,1) model. Therefore the GARCH(1,1) model is chosen. The GARCH(1,1) model is as follows.

$$\begin{aligned}
 INDEX_t &= c_t + \alpha_0 SENT_t + \varepsilon_t \\
 GARCH &= C(2) + C(3)RESID(1)^2 + C(4)GARCH(1)
 \end{aligned}
 \tag{5}$$

Table 20: T-GARCG(1,1)

Variable	Coefficient	Std. Error	z-Statistic	Probability
SENT	0.800918	0.018109	44.22869	0.00000
Variance Equation				
C	0.005059	0.002098	2.41175	0.0159
RESID(-1)^2	0.476706	0.166335	2.865941	0.0042
RESID(-1)^2*(RESID(-1)<0)	0.013539	0.196189	0.069012	0.945
GARCH(-1)	0.494433	0.068637	7.2036	0.0000

Table 21: E-GARCG(1,1)

Variable	Coefficient	Std. Error	z-Statistic	Probability
SENT	0.898765	0.016418	54.74405	0.00000
Variance Equation				
C(2)	-0.993838	0.240571	-4.131156	0.000000
C(3)	0.793765	0.171025	4.641216	0.000000
C(4)	-0.014632	0.087572	-0.167086	0.8673
C(5)	0.868212	0.064908	13.37613	0.000000

Figure 1 shows the transitions of the standardized INDEX and SENT. According to Figure 1, although there are differences in the fluctuation range with INDEX and SENT, the fluctuation trends of them are expected to be the same.

From all above, the effectiveness of SENT is be proved.

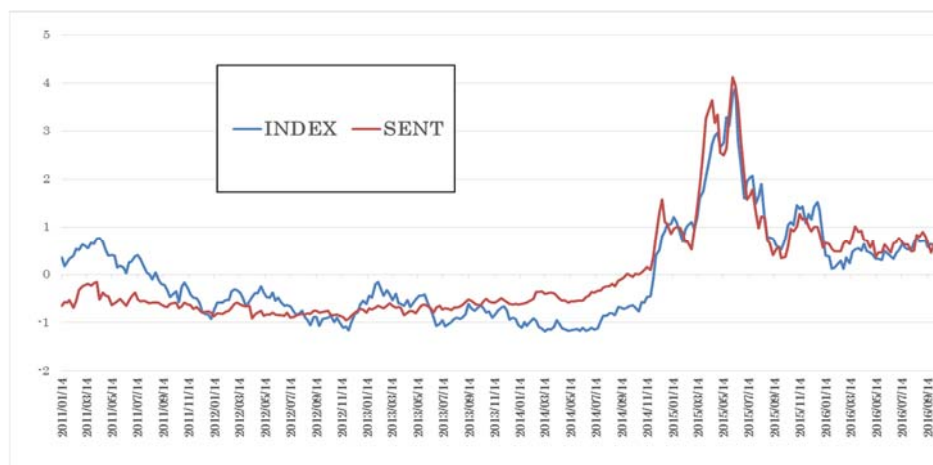


Figure 1: Transitions of INDEX and SENT

6.2 Consideration

The definition formula of the investor sentiment index proposed in this research is as follows.

$$SENT_t = -0.043L.CEFD + 0.167L.OPEN + 0.178TURN + 0.178MD \quad (2)$$

6.2.1 Consideration about lead-lag effect

According to equation (2), the lead-lag relationship between investor sentiment indices was as follows. The data before one term of CEFD and OPEN are considered to have a stronger influence on the investor sentiment than the current term. The reason for such a result is that there are many speculative investors in Chinese stock market, and the investor sentiment influenced by the resultant of these investors to make a time lag between the investor sentiment variables. In addition, it can be considered that the TURN and MD can directly reflect investor sentiment for the current term.

6.2.2 Consideration about the coefficient of the variables

According to the coefficient of the variables, they are consistent with previous work such as Baker and Wurgler [1], but the absolute value is seen to be relatively small. In other words, the influence of CEFD on investor sentiment is smaller than other investor sentiment indices. As for the reason why this phenomenon happens, it is that in the Chinese stock market, the closed-end funds are mainly held by institutional investors who are said to be relatively unlikely to be affected by investor sentiment. As for OPEN, TURN and MD, the influences of them on investor sentiment are considered to be similar.

6.2.3 Consideration about granger causality

According to the results of the granger causality test, investor sentiment affects the stock price index. To the contrary, stock price index also affects investor sentiment. Therefore, it seems that there is a circulation effect that expands the trend between stock price index and investor sentiment. The same effect is also discovered by Ma Ruwei and Zhang Na [6]. Such an action is thought to urge the formation of skyrocketing and collapse in stock price. This phenomenon is considered to have been seen especially in the period of mid-2015.

7 Conclusion

In this study, by using the BW method, we constructed a new investor sentiment index which is more compatible with the Chinese stock market. It also showed the investor sentiment index's effectiveness by comparing the stock price index. By analyzing the influence of investor sentiment on the Chinese stock market, it is also revealed that there is a circulation effect between investor sentiment and stock price index.

As for the future tasks, Baker and Wurgler [1] used macroeconomic indicators such as industrial production indices to make index that can reflect investor sentiment more purely by using multiple regression analysis. However, in this study, since we used weekly data for all stock price indicators, using monthly data of macroeconomic indicators will cause large errors. Therefore,

this research has omitted this part and devising an index synthesis method to build an index reflecting investor sentiment more purely.

Acknowledgement

I would like to thank my supervisor, Dr. Hasuike for his useful suggestions and support.

References

- [1] Malcolm Baker and Jeffrey Wurgler, "Investor Sentiment and the Cross-Section of Stock Returns," *The Journal of Finance*, Vol. 61, No. 4, pp. 1645-1680, 2006
- [2] Kenneth L. Fisher and Meir Statman, "Investor Sentiment and Stock Return," *Finance Analysts Journal*, Vol. 56, No. 2, pp. 16-23, 2000
- [3] Lee, Charles, Andrei Shleifer, and Richard H. Thaler, "Investor Sentiment and the Closed-end Fund Puzzle," *The Journal of Finance*, Vol.46, pp. 75-109, 1991
- [4] Robert Neal and Simon M. Wheatley, "Do Measures of Investor Sentiment Predict Returns?" *Journal of Financial and Quantitative Analysis*, Vol. 33, No. 12, pp. 523-547, 1998
- [5] Hu Changsheng and Chi Chunyang, "Investor Sentiment: Rational or Irrational," *Chinese Review of Financial Studies*, Vol. 6, pp. 46-62, 2012
- [6] Ma Ruowei and Zhang Na, "The Construction of Investor Sentiment Index for China's Stock Market: Based on the Panel Data of Shanghai A Share Companies" *Journal of Central University of Finance & Economics*, Vol. 7, pp. 42-49, 2015
- [7] Lu Xunfa and Lai Kinkeung, "The Relationship between Stock Indices and Investors' Sentiment Index in Chinese Financial Market" *Systems Engineering - Theory & Practice*, Vol. 32, No. 3, pp. 621-629, 2012