

The Construction of Investor Sentiment Index in China's Stock Market

Jiayu Gu

School of Economics and Management
Nanjing University of Science and Technology
Nanjing, China

Tao Li

School of Economics and Management
Nanjing University of Science and Technology
Nanjing, China

Abstract—Drawing on the method of Baker and Wurgler and combining with the national conditions of China, we select six indexes, such as the discount rate (CEFDR), the closed-end fund discount rate (CEFDR), IPO quantity (IPOV), price-earnings ratio (P/E), turnover rate (TRO) and consumer confidence index (CCI), which are the proxy indicators to measure investor sentiment changes, and a principal component analysis method is used to construct an index ISI that can better reflect the mood fluctuations of Chinese investors. In the process of using quantitative methods to study the effectiveness of investor sentiment index, this paper uses cointegration analysis and linear regression to verify it. The results show that ISI has a good ability to explain the volatility of China's stock market. The construction of ISI provides a good tool for the subsequent interpretation of investor sentiment hypothesis on the interpretation of "financial vision" in Chinese capital markets.

Keywords—investor sentiment; principal component analysis; index construction

I. INTRODUCTION

The measurement of investor sentiment has always been one of the key and difficult issues in the study of behavioral finance. The current research focuses on investor sentiment on stock returns volatility and market forecasts. However, the conclusions of different scholars on the same issue have shown quite different, and even diametrically opposite. The reason for the divergent conclusions of its research results is that, apart from different research methods and sample differences, it is likely that the accuracy of investor sentiment measurement is not high enough. According to the existing literature on investor sentiment measurement, the selected indicators can be divided into subjective indicators and objective indicators.

Brown and Cliff (2004) take investor intelligence index as the proxy indicator of investor sentiment, Lemmon and Portniaguins (2006) take consumer confidence index as the agent variable of investor sentiment, Wang Mei and Sun Jianjun (2004), Wang Zhen and Hao Gang (2014) adopt the number of new investors and consumer confidence index respectively to measure investor sentiment. The objective index is mainly based on the open trading information index in the stock market instead of investor sentiment. Ibbotson (1998), according to the closed-end fund discount rate,

Lawrence M Benveniste, Han Liyan (2007) based on the IPO circulation and the first day earnings, Baker and Stein (2004) according to the stock trading volume. At present, the measurement of investor sentiment at home and abroad is mostly based on a single substitute variable to measure directly or indirectly. This measurement method has the defects of single measurement dimension.

For this reason, some scholars try to combine multiple single indicators through various methods to form a comprehensive index to depict investor sentiment. Generally, investor sentiment index can reflect investor sentiment more comprehensively than single index. Baker and Wurgler (2006) select six objective market indicators, such as closed-end fund discount, market volume, IPO quantity, first day earnings and dividend premium, to build a comprehensive index to reflect investor sentiment. On this basis, Chinese scholars Yi Zhigao and maing (2009) constructed the investor sentiment index of China's stock market by using the consumer confidence index and the number of new investors' opening account reflecting the investor's subjective emotion. At present, the principal component analysis is used to construct the comprehensive index of investor sentiment. Baker and Wurgler (2006) use the first principal component, and Yi Zhigao and Maun (2009) use the weighted mean value of the first to five principal components to construct the emotional comprehensive index. This paper uses the research method of B-W emotional index and improves it, and integrates the indicators that can reflect the change of investor sentiment in China's stock market, and uses the principal component analysis to build a comprehensive index (ISI) to measure the investor sentiment of the Chinese stock market.

II. SELECTION OF VARIABLES AND ANALYSIS OF METHOD

A. Selection of Variables

Baker and Wurgler constructed a composite index to represent the emotional index based on 6 emotion proxy variables. The 6 proxy variables are the discount rate of closed-end funds, the exchange rate of the stock, the number of IPOs, the first day earn of IPOs, the proportion of newly issued bonds and stocks, dividend premium. Among them, the higher discount rate of closed-end funds indicates that

investors are not optimistic about the future market, so the discount rate of closed-end funds is negatively correlated with future earnings. Turnover rate characterizes market liquidity. High liquidity tends to be accompanied by overvaluation under the condition of market short selling limit and irrational investors, and turnover rate is also found to predict lower market returns. On the Chinese market, the first day of almost every stock IPO will touch the rise, that is, the first day is 44%, so the first day of IPOs will not be an effective proxy for the Chinese stock market. According to the characteristics of the Chinese stock market and the availability of the data, this paper finally selects six emotional proxy variables, such as CEFDR, TRO, IPOV, P/E, CCI, and new investor opening number (NIA), and use the principal component analysis to construct a comprehensive emotion index.

The above indicators are from January 2009 to November 2017, and they are monthly data. Fund discount rate (CEFDR), stock turnover rate (TRO), IPO quantity (IPOV) and price earnings ratio (P/E) are from Resset database and Sina Financial Website. The consumer confidence index comes from Orient fortune net. The number of new investors is from the central registration Clearing Corp.

It is necessary to note that there may be "advance" and "lag" in time when different indicators reflect investor sentiment. Jaffe (1975), Lowry and Schwert (2002) and Benveniste (2003), found that the first day yield of IPO was preceded by the number of IPO to reflect investor sentiment, that is, a large number of IPO events will appear before the first day of the high IPO yield. More generally, Baker and Wurgler (2006) think that the company's supply response variables are often lagging behind the variables that directly reflect investor needs (such as closed-end fund discount rates, turnover rates, trading volumes and the first day returns of the market). In view of this, this paper draws on the methods of Baker and Wurgler (2006), and selects 6 variables of current and advanced one.

B. Method Analysis

This paper constructs a comprehensive index of Chinese investor sentiment by principal component analysis. Since it is temporarily unable to determine which of the 6 current variables and 6 advance variables has greater impact on ISI, the following steps are taken for analysis. First, the correlation test of the 12 indicators is carried out to confirm whether the principal component analysis is applicable. Second, if there is a strong correlation between the 12 indexes, the principal component analysis is carried out for all the indexes, and the principal component is obtained. Third, using the main component obtained in the second steps to construct the Chinese investor mood comprehensive index, that is ISI₁, fourth, the total index ISI₁ is calculated. The current index and the early index, which have the highest correlation with index ISI₁ are selected. Fifth, the principal component analysis is carried out for the 6 indexes selected by the fourth step, and the principal component is obtained; sixth, the Chinese investor sentiment index (ISI) is constructed by using the main component of the fifth step.

III. THE CONSTRUCTION OF CHINESE INVESTOR SENTIMENT INDEX

A. Descriptive Statistics and Correlation Tests

"Table I" shows the descriptive statistics of the six variables. It can be seen that the number of new investors, the number of IPOs, and the turnover rate have greater fluctuations. The average number of newly added investor accounts was 1.156 million, and the standard deviation was 82.22.

TABLE I. DESCRIPTIVE STATISTICS

Variable Name	Mean	Standard Deviation	Sample Size
IPOV	17.76	14.25	107
CEFDR	0.02	0.10	107
TRO	32.1	16.27	107
CCI	104.9	5.01	107
NIA	115.6	82.22	107
P/E	25.8	4.97	107

a. Source: Resset, Orient Fortune Net, Sina Financial Website

Before the principal component analysis of the 12 variables, such as CEFDR, IPOV, P/E, TRO, NIA, CCI and their early indicators CEFDR-1, IPOV-1, TRO-1, P/E-1, CCI-1, NIA-1 and other variables, to ensure that the variables can be done well in the principal component analysis, the correlation test is carried out on the 12 variables. The test method is the test and Bartley. The test results are shown in "Table II".

TABLE II. KMO AND BARTLETT TEST

KMO and Bartlett test		
KMO sampling fitness number		.732
Bartlett's sphericity test	Approximate chi-square	1060
	Free degree	66
	Conspicuousness	.000

a. Source: Resset, Orient Fortune Net, Sina Financial Website

The KMO test coefficient is 0.732, and the results of Bartley spherical test show that Sig is 0.00. The two methods all indicate that the correlation of each index is higher, so the 12 variables are suitable for the principal component analysis.

B. Building an Emotional Index ISI₁

After determining the correlation of variables, 12 variables are normalized and principal component analysis is done. "Table II" shows that the cumulative variance contribution of the first four principal components is 81.56%. Therefore, the Chinese investor sentiment index ISI₁ was selected from the first four principal components to construct 12 variables. It should be noted that, in the course of calculation, the method of constructing B-W emotion index is improved, that is, this paper strictly observe the statistical standard of at least 80% of the cumulative variance interpretation rate, and use the weighted average of first, second, third and 4 principal components each time. The weight of each principal component is the variance contribution rate, so that more information can be retained.

TABLE III. EXPLANATION OF TOTAL VARIANCE

ingredient	variance percentage	Cumulative (%)
1	40.825	40.825
2	25.202	66.027
3	8.308	74.335
4	7.233	81.568

a. Source: Resset, Orient Fortune Net, Sina Financial Website

According to the initial factor load matrix, we find that the first principal component mainly reflects the basic information of the variables of NIA, P/E₋₁, TRO, P/E; second main components mainly reflect the information of CCI₁, CCI, IPOV₋₁; the third principal component reflects the information of the variables such as CEFDR, IPOV, and the fourth principal component reflects the information of the variables CEFDR₋₁, IPOV₋₁.

TABLE IV. COMPONENT MATRIX

	ingredient			
	1	2	3	4
CEFDR	.182	-.048	.869	-.089
TRO	.835	-.105	.345	.179
CCI	.036	.936	-.020	-.091
NIA	.831	.222	.001	-.031
P/E	.895	.100	-.047	.009
IPOV	.309	.569	-.476	-.322
IPOV-1	.257	.595	-.392	-.389
CEFDR-1	.177	-.129	-.061	.901
TRO-1	.841	-.113	.169	.293
CCI-1	.053	.960	-.001	.008
NIA-1	.824	.220	-.089	-.018
P/E-1	.917	.112	-.035	-.078

a. Source: Resset, Orient Fortune Net, Sina Financial Website

Therefore, the four principal components extracted can basically reflect the information of all variables, and these four new variables can be used instead of the original variables. After calculating the four principal component expressions, we can get the Chinese investor sentiment composite index of 1.

$$ISI_1 = 0.387 * F_1 + 0.219 * F_2 + 0.108 * F_3 + 0.101 * F_4 \quad (1)$$

C. Correlation Analysis Between ISI_1 and 12 Variables

This paper successfully constructed a comprehensive index of China's investor sentiment ISI_1 , by using principal component analysis. The index is a function of the advance and the current variable of the 6 index, because the index contains the current value and the advance value of the same indicator, it is necessary to make a trade-off between the current value and the advance value. This paper calculates the Pearson correlation coefficient of 12 variables and ISI_1 , and elects 6 variables with larger correlation coefficient as the component index of the Chinese investor sentiment index in the next stage. As shown in Table 5, the correlation coefficients of CEFDR and CEFDR₋₁ with ISI_1 are -0.390 and -0.266 respectively, indicating that the discount rate of closed-end funds is inversely proportional to the Chinese investor sentiment index, that is, the greater the CEFDR (the higher the discount rate), the lower the investor sentiment (the smaller), this conclusion is consistent with the results of the Baker and Wurgler (2006). The relative coefficients of

IPOV and IPOV₋₁ with ISI_1 are 0.342 and 0.316 respectively, indicating that the number of the investors is positively related to the investor sentiment, that is, the more the IPO events, the higher the investor sentiment; the correlation coefficient of the P/E and P/E₋₁ with ISI_1 is 0.784 and 0.771 respectively, indicating that the first day return of the market is positively correlated with the investor sentiment, and the correlation coefficient of TRO and TRO₋₁ the correlation is , respectively 0.743 and 0.749, indicating that the turnover rate is positively related to investor sentiment, and both are significant at the level of 1% significant, that is, the higher the turnover rate, the higher the investor sentiment; the correlation coefficient of NIA₋₁ and NIA, the correlation is 0.749 and 0.784, respectively, indicating that the number of new investors is proportional to the investor sentiment, and the CCI and CCI₋₁ the correlation coefficient with ISI_1 are 0.422, and 0.494 respectively, it shows that the consumer confidence index is positively related to investor sentiment, that is, the higher the consumer confidence index, the higher the investor's mood, which can be explained from the wealth effect of the securities market. When the stock market is good, the consumer expects wealth to increase, and the confidence of the consumer is enhanced. In addition, this paper also compares the early and current variables of the 6 indexes, and finds that the correlation coefficient of the current variables of TRO and CCI with ISI_1 is lower than the one in advance, that is, the market turnover rate reflects the investor sentiment in advance, and the confidence index reflects the expected behavior of the consumer, so it also reflects the investor sentiment in advance. The remaining four indexes are completely opposite. The correlation coefficient of IPOV is larger than that of IPOV₋₁, which is consistent with the results of Baker and Wurgler (2006). Finally, this paper selects 6 variables, CEFDR, IPOV, NIA, P/E, CCI₁, TRO₋₁ and so on to construct the Chinese investor sentiment comprehensive index (ISI).

After the above analysis, the paper finally selects 6 variables: CEFDR, IPOV, P/E, CCI₁, TRO₋₁ and so on to construct the Chinese investor sentiment comprehensive index ISI. The method of principal component analysis is still used in the specific method, and the expression of China's investor sentiment index is:

$$ISI = -0.021CEFDR + 0.153IPOV + 0.156P/E + 0.095TRO_{-1} + 0.165NIA + 0.125CCI_1 \quad (2)$$

D. Analysis of Consistency Between ISI_1 and ISI

By comparing the investor sentiment index ISI_1 and the investor sentiment index ISI, we can find that the correlation between the two investor sentiment indices is up to 86.4% and is significant at the 1% level, and the following figure shows a similar conclusion. It shows that this paper deletes 6 investors' emotional proxy indicators that have little influence on the construction of investor sentiment index, that is, the investor sentiment index ISI can better reflect the change of investor sentiment.

TABLE V. CORRELATION COEFFICIENT BETWEEN ISI1 AND 12 INDICATORS

	ISI-1	IPO-1	CEF-1	TRO-1	CCI-1	NIA-1	P/E-1	IPO	CEF	TRO	CCI	NIA	P/E
ISI-1	1												
IPO-1	.316**	1											
CEF-1	-.26**	-.221*	1										
TRO-1	.749**	-.043	.352**	1									
CCI-1	.494**	.512**	-.165	-.053	1								
NIA-1	.759**	.418**	.090	.727**	.247*	1							
P/E-1	.771**	.083	.334**	.733**	.169	.721**	1						
IPO	.342**	.732**	-.169	-.016	.527**	.377**	.362**	1					
CEF	-.31**	-.138	.064	.242*	-.103	.075	.142	-.204*	1				
TRO	.743**	-.086	.255**	.779**	-.031	.567**	.674**	-.046	.356**	1			
CCI	.442**	.538**	-.27**	-.068	.876**	.227*	.167	.49**	-.143	-.076	1		
NIA	.784**	.295**	.085	.557**	.262**	.705**	.684**	.41**	.110	.730**	.219*	1	
P/E	.784**	.254**	.146	.679**	.156	.608**	.901**	.34**	.089	.729**	.166	.722**	1

a. Source: Resset, Orient Fortune Net, Sina Financial Website

TABLE VI. CORRELATION ANALYSIS

		ISI1	ISI
ISI1	Pearson correlation	1	.864**
	conspicuousness		.000
ISI	Pearson correlation	.864**	1
	conspicuousness	.000	

a. Source: Resset, Orient Fortune Net, Sina Financial Website

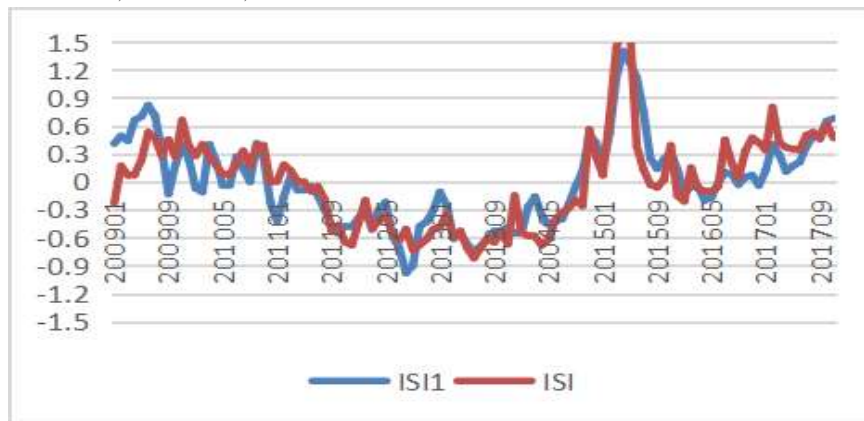


Fig. 1. The trend of ISI and ISI₁.

Compared with the investor sentiment metrics previously studied, the ISI constructed in this paper contains six variables directly related to investor sentiment. Compared with the previous study of investor sentiment metrics, the ISI constructed in this paper contains six variables directly related to investor sentiment, and takes into account the current value of each variable and the value in advance. Compared with a single metric, the information content is higher.

IV. ROBUSTNESS CHECK

The investor sentiment index ISI is built in this paper. Generally speaking, investor sentiment should be consistent with the trend of the stock market. When the market is better, investor sentiment should go up and vice versa. This article will test whether the investor sentiment index has this feature.

A. Qualitative Analysis of the Consistency Between ISI and the Trend of Chinese Stock Market

"Fig. 2" is the change trend of the monthly return rate of Shanghai Composite Index and ISI. This index fits well the fluctuation of Chinese stock market during the period of the sample, which shows that the ISI in this article is better able to explain the volatility of the stock market in China.

TABLE VII. CORRELATION ANALYSIS BETWEEN ISI AND ISI WITH R

		ISI1	ISI	r
ISI1	Pearson correlation	1	.864**	.313**
	conspicuousness		.000	.001
ISI	Pearson correlation	.864**	1	.289**
	conspicuousness	.000		.003
r	Pearson correlation	.313**	.289**	1
	conspicuousness	.001	.003	

a. Source: Resset, Orient Fortune Net, Sina Financial Website

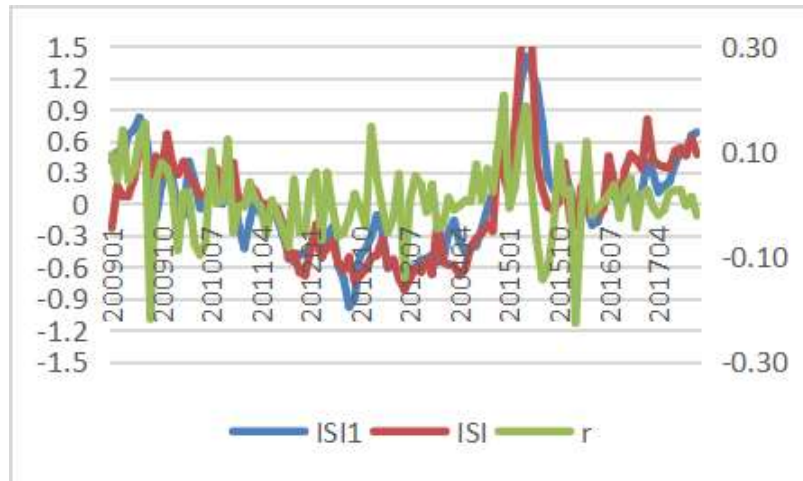


Fig. 2. The trend of investor sentiment index and r.

In order to cushion the negative impact of the international financial crisis on China's economy, the Chinese government implemented the four trillion investment plan from late 2008 to August 2009 to stimulate the development of the domestic economy. From September 2009 to mid-2014, China's overall economic environment is poor, overcapacity, and the economic recovery is slow. As a result, the Shanghai Composite Index fluctuated and the

stock market suffered another bear market. In this period, the investor sentiment index has fluctuated in the short term, but the overall downward trend is in the downward trend. The index characteristics of the national stock market are basically consistent. From mid-2014 to June 2015, China's stock market went through another bull market. Meanwhile, investor sentiment index continued to rise during the same period, and then fell to a trough.

TABLE VIII. UNIT ROOT TEST

	ADF	1%	5%		Inspection form	
variable	statistical quantity	critical value	critical value	P value	Lag order	conclusion
ISI	-2.66	-2.58	-1.94	0.0081	0	steady
r	-8.92	-2.58	-1.94	0.0000	0	steady

a. Source: Resset, Orient Fortune Net

B. Quantitative Analysis on the Effectiveness of Investor Sentiment Index

In the process of using quantitative methods to study the effectiveness of investor sentiment index, this paper uses cointegration analysis to verify it. Through cointegration analysis, we discuss whether there is a long-term equilibrium relationship between investor sentiment index and stock index.

Both ISI and r are stationary.

"Table VII" shows the cointegration test between the investor sentiment index (ISI) and the Shanghai Composite Index, which indicates that there is a cointegration relationship between the investor sentiment index (ISI) and the monthly yield of the Shanghai Composite Index (SSE), that is, there is a long-term equilibrium relationship.

The following is a review of the current dependence of investor sentiment indices on market returns.

$$r_t = \alpha + \beta \times ISI_t + \varepsilon_t \quad (3)$$

TABLE IX. COINTEGRATION TEST

null hypothesis	characteristic root	Trace check quantity	5% threshold	P value
No cointegration relation	0.53	83.57	15.49	0.000
at most one cointegration	0.05	5.55	3.84	0.018

a. Source: Resset, Orient Fortune Net, Sina Financial Website

TABLE X. REGRESSION RESULTS OF INVESTOR SENTIMENT INDEX

Emotional index	β	P value	T value	R ² (%)
ISI	0.042	0.0026	3.089	8.33

b. Source: Resset, Orient Fortune Net, Sina Financial Website

The coefficient of ISI is significant at the level of 5%, indicating that the constructed emotional index is dependent on the yield of the current period.

V. CONCLUSION

This article uses Baker and Wurgler's (2006) method to combine China's national conditions and select six indicators such as closed-end fund discount rate, IPO quantity, IPO first day yield, market trading volume, turnover rate and consumer confidence index as the proxy index to measure investor sentiment changes, and constructs an index ISI that can better reflect the mood fluctuations of Chinese investors.

In the process of using quantitative methods to study the effectiveness of investor sentiment index, this paper mainly uses cointegration analysis to verify. The results of the study indicate that ISI has a better ability to explain the volatility of the Chinese stock market. The construction of the ISI provides us with a good tool for the subsequent study of the investor sentiment hypothesis on the interpretation of the "financial vision" of the Chinese capital market.

REFERENCES

- [1] Brown, Gregory W and Michael T Cliff. Investor Sentiment and the Near-term Stock Market[J]. *Journal of Empirical Finance*, 2004(11):1-27.
- [2] Lemmon, M and Portniaguina, E. Consumer Confidence and Asset prices: Some Empirical Evidence[J]. *Review of Financial Studies*, 2006(19): 1499-1529.
- [3] Ibbotson R G. Price performance of common stock new issues[J].*Journal of Financial Economics*,1975 (2) :235 -272.
- [4] Lawrence M Benveniste. Book building vs Fixed Price: An Analysis of Competing Strategies for Marketing IPOs[J].*The Journal of Financial and Quantitative Analysis*,1997,4 (32) : 383 -403.
- [5] Baker,Wurgler.Investor sentiment and the cross-section of stock return[J].*Journal of Finance*, 2006 (1) :1645 -1680
- [6] Wang Meijin, Sun Jianjun. China's stock market returns, earnings volatility and investor sentiment [J]. *Economic Research*, 2004, (10): 75-83.
- [7] Cheng Kun, Liu Renhe.Investigation of investor sentiment and stock market[J].*Shanghai Economic Research*,2005,(11):86-93.
- [8] Han Liyan, Wu Yanran. The Enigma of Investor Sentiment and IPOs - Underpricing or Premium[J].*Management World*, 2007(3): 51-61.
- [9] Yi Zhi Gao, Mao Ning. Research on Investor Sentiment Measurement in Chinese Stock Market[J].*Finance Research*,2009(11).
- [10] Xu Haichuan, Zhou Weixing. Emotional Index and Market Returns: An Analysis of the Incorporation of China's Hangzhi Index[J]. *Journal of Management Science*, 2018(1): 88-96.