Construction of Chinese Stock Investor Sentiment Index

Ying Zhao(✉)

Johns Hopkins University, Baltimore, USA
zhaoyingalex@outlook.com

Abstract. Based on some data of the CSI 300 index, this paper uses the principal component analysis (PCA) to build the investor sentiment index, and then discusses the feasibility of constructing the investor sentiment index. It aims to build an investor sentiment index that matches the Chinese market, understand the psychological characteristics of investors, predict their expectations, and formulate effective regulatory policies.

Keywords: stock market · investor sentiment · stock yield · sentiment index construction

1 Introduction

Economists have long tried to measure investor sentiment because it is the best indicator of their willingness to buy or sell products in the current market [1]. There is no denying that every investor is faced with this sentiment, but because different investors have different economic levels and positions, this index often gives different answers. At present, the mechanism of China's bond market is not yet perfect, so the sentiment index analysis is closer to the actual situation of investors exploring the market situation, which is conducive to explaining various financial anomalies [2]. Under the guidance of behavioral finance theory, investor sentiment can be used to analyze China's stock price factors more comprehensively and objectively, and provide theoretical and empirical support for Chinese investors to undertake the reality, which is very important for the regulatory policies of investors and governments. This paper takes the theory of investor sentiment in behavioral finance as the research theory, refers to the current research situation at home and abroad, adopts the principal component analysis method to extract the principal components, and finally constructs an index that can reflect the investor sentiment.

2 The Influence Mechanism of Investor Sentiment

Behavioral economics is a term for emotional heuristic bias that means that people may not like making correct decisions through rational logical analysis and factual judgment, but it often the user experience or emotion-driven decisions [3]. Due to the fast and simple

F. Balli et al. (Eds.): ESFCT 2022, AEBMR 226, pp. 105–112, 2022.
https://doi.org/10.2991/978-94-6463-052-7_13

decision-making, this method is often used for decision-making behavior. Therefore, it is important for investors to have more direct positive emotions when investing, due to a heuristic emotional bias. If investors can have more positive sentiment, then they are likely to be more optimistic about the future [4]. Optimists, when depressed, are more likely to make positive investment decisions than pessimists, and are more likely to ignore the bad news lurking in the market.

3 The Construction Method of Investor Sentiment Index

3.1 The Measurement Method of the Investor Sentiment Index

This paper adopts the principal component analysis (PCA) to create emotional index model, the principle of the PCA method is through a variety of calculation methods will be multiple indicators into several new indicators, new index construction characteristic is not significant and the mean variance contribution rate is declining, and the first element of the corresponding variance contribution rate is the largest, thus eliminating the multiple collinearity in the model [5].

3.2 Selection of Investor Sentiment Index Variables

Based on the existing research results, based on data availability and comprehensive investigation, this paper decides to use the China Consumer Confidence Index (CCI) as the subjective sentiment index, p/E (PE), rise or fall (ZDF), trading volume (CJL) and the number of new investor accounts opened (NIA) as the sentiment index [6–8].

3.3 Selection and Processing of Index Data

The data of this article are from Oriental Fortune network and China Securities Registration Corporation data Center [9]. The latest authoritative monthly monthly data for the five years from February 2016 to January 2021 is the weighted average of each stock, and the p/E ratio adopts the static weighted average p/E ratio.

3.4 Descriptive Statistics of the Data

After collecting and processing the data of each index, the following descriptive statistical analysis was obtained:

The results of the statistical analysis are presented in Table 1 and Table 2. From February 2016 to January 2021, there were 60 sets of monthly data, among which the maximum CCI was 127.6, the minimum was 99.8, and the average was 118.35, indicating that the overall value is relatively stable, indicating that with the rapid development of Chinese economy, Chinese consumers are unanimously optimistic about the future economy and relatively optimistic. The highest PE was 48.71%, the lowest was 25.6%, and the average was 35.9%, with a large deviation, and significant fluctuations, indicating that the p/E ratio can reflect the relationship between the stock price and the current price. In general, a low p/E ratio means that the CJL stock is close to its current price. In

Table 1. Raw data required to construct investor sentiment indicators

Time	CCI	PE (%)	CJL (hand)	ZDF (%)	NIA (Thousandsoffamilies)
2016/2	104.4	35.95%	10664778169	−12.50%	115.5
2016/3	100	36.84%	13319891643	9.75%	229.2
2016/4	101	35.94%	10304436770	−9.25%	155.1
2016/5	99.8	35.62%	7550473800	2.59%	142.1
2016/6	102.9	38.94%	8434378035	−2.01%	142.1
2016/7	106.8	38.95%	11053162814	7.81%	130.6
2016/8	105.6	39.91%	10316718974	6.86%	219.9
2016/9	104.6	38.51%	7204680355	−11.07%	185.4
2016/10	107.2	38.79%	9274721969	5.93%	131
2016/11	108.6	39.26%	14485249986	6.88%	174.8
2016/12	108.4	40.12%	11529803350	−9.91%	139.7

(Only the data of 2016 are displayed for reference)

Table 2. Descriptive analysis results of different mood indicators

Indicators		CCI	PE (%)	CJL (hand)	ZDF (%)	NIA (Thousandsoffamilies)
The case number	effective	60	60	60	60	60
	missing	0	0	0	0	0
The average		118.35	35.90%	11800266452.25	1.46%	132.36
The median		121.90	37.02%	10657396473.01	2.72%	122.40
The standard deviation		8.15	5.14%	3978462179.92	3.31%	41.81
The minimum value		99.80	25.60%	6998046373.68	−13.27%	74.24
The maximum		127.60	48.71%	25853130291.30	13.05%	242.60

behavioral finance, stock prices are directly influenced by investor sentiment. Thus, after 2016, investors' investment sentiment fluctuated greatly; the CJL extreme value changed greatly, but still at a large amount. It can be speculated that China's stock market has been active since 2016, and the market momentum is strong in the future. We use index change (ZDF) to represent activity across the stock market. It can be seen that the rise or fall has not changed much, but the range is large, indicating that the stock market will show a high or even unrealistic active, but there will also be a very large decline. The NIA maximum varies widely, reflecting possible changes in investor sentiment outside the market. When the market is good, there will be a large number of investors pouring

Table 3. KMO and Bartlett tests

KMO sampling suitability quantity		.643
Bartlett sphericity test	The approximate chi-square	403.156
	Degrees of freedom	45
	Significant	.000

into the investment market, and when the market is not good, most choose to wait and see [10]. But the average annual attendance of 1.32 million people shows that the Chinese stock market is very active.

3.5 Build Up Investor Sentiment

This paper will use SPSS software to obtain the CSI 300 index data based on PCA (principal component analysis) [11], Establish an investor sentiment index model and conduct a feasibility study.

First, the KMO test and the sphericity Bartlett test were used to determine whether five indicators and five first-order lag indicators (represented in X, see Table 3) were applicable to principal component analysis [12, 13].

The KMO test value was $0.652 > 0.5$, and the P-value was much less than 0.05, indicating that the data were suitable for the principal component analysis.

As can be seen from Table 4, the initial eigenvalues of the first three components are all greater than 1, and the cumulative contribution rate of the first three components can reach 72.17%, which can be basically explained overall. Therefore, the first three components were extracted.

Based on the coefficients in the component matrix and the eigenvalues of the first three components, given by the formula

$$u_{ij} = \frac{r_{F_i X_i}}{\sqrt{\lambda_i}} \tag{1}$$

The expression of the principal components of the model is:

$$\begin{aligned} F1 = &-0.266CCI - 0.229CCIx + 0.405PE + 0.404PEx + 0.289CJL \\ &+ 0.276CJLx + 0.256ZDF + 0.175ZDFx + 0.364NIA + 0.401NIAx \end{aligned} \tag{2}$$

$$\begin{aligned} F2 = &0.532CCL + 0.553CCLx + 0.018PE - 0.005PEx + 0.371CJL \\ &+ 0.399CJLx + 0.203ZDF + 0.244ZDFx - 0.105NIA - 0.028NIAx \end{aligned} \tag{3}$$

$$\begin{aligned} F3 = &0.077CCI + 0.184CCIx + 0.440PE + 0.352PEx - 0.129CJL \\ &- 0.277CJLx - 0.5553ZDF + 0.461ZDFx - 0.057NIA - 0.154NIA \end{aligned} \tag{4}$$

$$II = 0.519F1 + 0.323F2 + 0.158F3 \tag{5}$$

You can construct the initial model as:

$$II = 0.046CCI_t + 0.089CCI_{t-1} + 0.285PE_t + 0.264PE_{t-1} + 0.25CJL_t + 0.228CJL_{t-1} + 0.111ZDF_t + 0.243ZDF_{t-1} + 0.146NIA_t + 0.175NIA_{t-1} \quad (6)$$

Table 4. Eigenvalues and variance contribution rates of each principal component

composition	Initial eigenvalue			Extract the sum of squares of loads		
	A total of	Percentage of variance	Cumulative %	A total of	Percentage of variance	Cumulative %
1	3.746	37.464	37.464	3.746	37.464	37.464
2	2.331	23.307	60.771	2.331	23.307	60.771
3	1.140	11.400	72.171	1.140	11.400	72.171
4	.929	9.293	81.464			
5	.704	7.036	88.500			
6	.549	5.493	93.993			
7	.271	2.712	96.705			
8	.216	2.160	98.864			
9	.070	.695	99.560			
10	.044	.440	100.000			

(Extraction method: principal component analysis)

Table 5. II correlation analysis with each variable

	CCI	CCIx	PE (%)	PE x	CJL (hand)	CJL x (hand)	DZF (%)	DZFx	NIA (Thousands of families)	NIAx (Thousands of families)
II	−.137	−.112	.482	.548	.746	.823	.564	.374	.548	.738
	.298	.396	.000	.000	.000	.000	.000	.000	.000	.000

[a]The column N = 60

Table 6. KMO and Bartlett tests for new proxy variables

KMO sampling suitability quantity		.704
Bartlett sphericity test	The approximate chi-square	49.000
	Degrees of freedom	6
	significant	.000

After obtaining the initial model of the emotion index, the model results were associated with the previous variables, and the index with the highest correlation with II was selected as the final variable.

As the Table 5, the correlation between CCI, CCIx, PE, CJL, DZFx, NIA, and II is relatively small. Therefore, the cancellation of these six metrics was considered, and PEx, CJLx, DZF, and NIAx were selected as the final proxy indicators. Reconcculating the new proxy using PCA yields the following results:

It can be found that the KMO value is 0.704, much larger than the KMO value obtained from the original index, indicating that the test variable is more suitable for the principal component analysis, and has passed the dominance test (Table 6).

Repeat the above steps to obtain the final pattern of:

$$II = 0.455PE_{t-1} + 0.555CJL_{t-1} + 0.47ZDF_t + 0.514NIA_{t-1} \tag{7}$$

This formula is an investor sentiment index model established based on the CSI 300 index, through which investor sentiment can be quantified to explore the quantitative relationship of investor behavior and better understand the dynamic changes of the market [14].

4 Model Effectiveness Analysis

Since the investor sentiment model established in this paper aims to guide investors to invest for satisfactory returns, the relationship between the investor sentiment model and the return is the key to test whether the model is effective. This paper uses the monthly yield of the CSI 300 index and the sentiment index model constructed for the correlation analysis. The results are shown in Table 7.

It can be seen that the correlation coefficient of the two indices was 0.696 and a p-value < 0.01. It can be seen that the sentiment index model constructed in this paper has a strong correlation with the income of CSI 300 index, and further analyzes the images corresponding to the monthly return of CSI 300 index and the sentiment index (see Fig. 1): The image of the sentiment index constructed in this paper is somewhat consistent with the yield curve trend of CSI 300 index in the same period. It can be seen

Table 7. CSI 300 yield and sentiment index correlation

			ii	syl
szpilmanRho	ii	The correlation coefficient	1.000	.696**
		Sig.(Double tail)	.	.000
		N	60	60
	syl	The correlation coefficient	.696**	1.000
		Sig.(Double tail)	.000	.
		N	60	60

**At 0.01 level (two-tailed), the correlation was significant

Table 8. Granger causality test of sentiment index II and CSI 300 index returns

Pairwise Granger Causality Tests			
Date:04/15/21Time:22: 31			
Sample:2016M022021M01			
Lags:2			
NullHypothesis:	Obs	F-Statistic	Prob.
SYLdoesnotGrangerCauseII	58	9.75950	0.0004
IIdoesnotGrangerCauseSYL		10.2735	0.0002

that the sentiment index model adopted in this paper can well adapt to the change of market returns and has some rationality.

To further analyze its correlation with the investment market, we conducted a Granger causal test for the returns of the sentiment index II and the HS300 index (see Table 8). We found a Granger causal relationship between the mood index and the HS300 index return, which is consistent with the findings of other researchers in China [15]. Through the causal test results, it can be seen that the investment situation index constructed in this paper is closely related to the market and forms a causal relationship, so as to effectively predict the fluctuation trend of the stock market through the constructed investor sentiment index and show its potential value and significance.

5 Conclusion

As the economy advances, China's stock market will prosper more and more, but it will start late and cannot have as strong market forces as Western countries have. The Chinese stock market has been favored by retail investors compared with mature investment institutions in the West. The impact is the biggest, so for the Chinese people, personal investment is all for research, and investor sentiment is at the heart of the problem. Based on the previous research on the investor sentiment index, this paper selects 5 indicators (CCI, the number of new investors per month, the PE p/E ratio of the CSI 300 index, the trading volume CJL and the rise and fall ZDF), and adopts the PCA principal component analysis method to analyze and model it and its lagging data. Then the optimal agent variable is selected, and the model is built again using the PCA method to obtain the final investor sentiment index model (II).It can not only help investors to predict the future trend, understand the current stock trend, and make theoretical guidance to the stock practice. It can also help market regulators to understand the rules of market operation, understand the psychological characteristics of investors, predict their expectations, and formulate effective regulatory policies.

References

1. Yu X Y. The Construction and empirical Research of the Stock Market Investor Sentiment Index in China [D]. Henan University of Economics and Law, 2020.

2. Freybote J, Seagraves P. The Impact of Investor Sentiment on Commercial Real Estate Market Liquidity[J].Journal of Real Estate Research, 2018, 40(4):597–627.
3. river source. Construction and Effectiveness Test of Investor sentiment Index in China [J]. Market Research, 2020 (02): 17–19.
4. He L T, Casey K M. Improvements in forecasting of bank stock excess returns using the investor sentiment endurance index: a comparison with CAPM and Fama-French models [J]. International Journal of Financial Markets & Derivatives, 2018, 6(3):210–224.
5. Zhang Z C, concubine Liu Yuchen. Construction and Effectiveness Test of Investor sentiment Index in China [J]. Business Economy Research, 2018 (07): 156–158.
6. Lai C. Investor sentiment by relative strength index and stock return: Empirical evidence on Vietnam's stock market [J]. Accounting, 2021:451–456.
7. Li Y State. Research on Construction of Investor Sentiment Index and Its Impact on Stock Market Forec Effect [D]. Yunnan University of Finance and Economics, 2020.
8. Yu H. Research on the Systemic Financial Risk of China's Stock Market Based on the Dynamic Correlation of Industry Sectors [D]. Shandong University of Finance and Economics, 2022.
9. Alrabiah A. INVESTOR SENTIMENT AND STOCK INDEX RETURNS: THE CASE OF SAUDI STOCK EXCHANGE (TADAWUL) [J]. Journal of International Finance and Economics, 2019, 19(2):25–34.
10. Chen W, Zhou S, Zhang Y, et al. A Study of Multi-Scale Relationship Between Investor Sentiment and Stock Index Fluctuation Based on the Analysis of BEMD Spillover Index[J]. Journal of Systems Science and Information, 2021,9 (4): 399–420.
11. Deng C Z, Wang M. A Study on the Impact of Investor Emotions on Stock Market Earnings [J]. Market Week, 2021,34 (09): 110–112.
12. Zi-Long L, Su-S W, Ming-Zhu H. International investor sentiment and stock returns: Evidence from China [J]. Investment Analysts Journal, 2021, 50(12):1–17.
13. Reis P, Pinho C."A dynamic factor model applied to investor sentiment in the European context" [J]. Investment Management and Financial Innovations, 2021, 18(1):299–314.
14. Tian L. Construction of investor sentiment indicators and its impact on stock market earnings [J]. Chinese Business Theory, 2019 (13): 64–65.
15. He Z X. Study on the Influence of Investor Emotions on China Stock Market Income [D]. Yunnan University of Finance and Economics, 2022.