

# Investor Sentiment Measurement and Time Series Analysis

Yuchen Bai

School of Finance, Shanghai University of International Business and Economics, Shanghai, 201100, China

**Abstract.** Investor sentiment is one of the destabilizing factors in the stock market. In the past, some scholars have proposed models for the measurement of emotions. On this basis, this paper selects 9 factors (closed-end fund discount rate, trading volume of the previous month, market turnover rate of the previous month, number of IPOs, first-day yield of IPOs, number of new investors opened accounts in the last month, Consumers Confidence index, Advance/Decline Line, Business Index of Macro-economic), and constructs the investor sentiment index through principal component analysis. Through statistical analysis of the index, it is found that it satisfies the conditions of a stationary time series, and a related formula is put forward. At the same time, the relationship between index and stock market is discussed.

## 1. Introduction

In economics, there is a rational person hypothesis, which believes that the economic actions taken by every person engaged in economic activities are to obtain the greatest benefits at the least cost. In asset pricing models, there are similar assumptions. The capital asset pricing model includes the assumption that investors obey the Dominance rule, that is, under the same risk level, they choose securities with higher yields. Arbitrage Pricing Theory also believes that investors are risk averse and want to maximize utility.

But in reality, investors are not completely rational, and their behavior can be influenced by emotions or preferences, causing asset prices to deviate from their value. Behavioral finance argues that people tend to place too much trust in their own judgments. Irrationality when investing seems to be a common problem and has always been there. For example, in the face of loss, people generally have loss aversion. This is the psychology that a group will have, and it cannot be hedged by the number of people. Therefore, investor sentiment should also be regarded as a systematic risk. Therefore, investor sentiment has an impact on asset pricing and market stability, especially when China's stock market and participants' ideas are immature.

Based on China's A-share market, this paper measures consumer sentiment and studies its statistical significance. The second part of the article reviews the existing research at home and abroad. The third part is the choice of research methods and data. The fourth part is the construction of the index. The fifth part is statistical analysis. The sixth part is the relationship between index and stock price. The last part is the conclusion.

## 2. Literature review

Investor sentiment refers to the systematic deviations investors have towards future expectations. This concept was first proposed by Stein in 1996. For the calculation of investor sentiment, there is no uniform standard. The most popular view is that the number stands for the overall optimistic or pessimistic judgment of investors on stocks. On the one hand, the research on investor sentiment includes foreign calculations and verifications, and on the other hand, domestic scholars have carried out effective research and exploration on the development of this concept in terms of China's stock market.

Baker M and Wurgler J (2004) challenged classical finance theory and proposed that investor sentiment had a significant cross-sectional effect.[1] They found that stocks with small, high volatility, unprofitable, non-dividend-paying characteristics had lower subsequent returns when market sentiment was high. Achim Klein, Martin Riekert, Lyubomir Kirilov and Joerg Leukel (2018) analyzed the relationship between investor sentiment from news and stock prices. [2]

Zhigao Yi and Ning Mao (2009) constructed a comprehensive index that could better measure investor sentiment in China's stock market while controlling for the impact of economic fundamentals on sentiment, based on individual sentiment indicators such as the discount of closed-end funds, the number of IPOs and the first day of listing, the consumer confidence index and the number of new investors opening accounts. [3] Xingji Wei, Weili Xia and Dandan Sun (2014) selected 6 sentiment variables such as market turnover rate, and used the principal component analysis method to construct a monthly investor sentiment index in China's securities market. [4] They found that the constructed sentiment index could accurately measure the general sentiment of investors.

---

20013007@suibe.edu.cn

However, the construction of investor sentiment index can continue to add useful factors and improve the effectiveness of the index. This paper assumes that investor sentiment and the stock market are mutually influenced. The rise and fall of stock prices will drive investor sentiment and cause instability. Changes in investor sentiment can change investors' decision-making, which further contributes to the volatility of stock prices. At the same time, this paper will also try to analyze and predict the index from a statistical point of view.

### 3. Research methods and data

#### 3.1. Principal Component Analysis

Principal component analysis is to recombine many of the original factors with certain correlation into a new set of comprehensive indicators that are independent of each other to replace the original indicators. This paper refers to the construction method of Investor Sentiment Index [3] and CICI [4], and selects the closed-end fund discount rate, the trading volume of the previous month, the market turnover rate of the previous month, the number of IPOs, the first-day yield of IPOs, the number of new investors opened accounts in the last month, Consumers Confidence index, Advance/Decline Line, and Business Index of Macro-economic as factors. The data selects the monthly data from January 2011 to May 2022 and comes from CSMARSolution. SPSS is chosen as the data analysis software.

#### 3.2. Factors selected

**Closed-end Fund Discount.** The mystery of closed-end funds refers to a financial phenomenon in which the price and net asset value of closed-end funds are in a state of deviation for a long time. One of the reasons for the fund discount is investor sentiment. The essence of the fund is also a basket of stocks. When investors expect that the closed funds they invest in are more risky than the fund's portfolio, the price they are willing to pay will be less than the fund's NAV, resulting in a fund's discount. Investors' sentiment can influence their judgment of risks. So the closed-end fund discount rate is also used to represent investor sentiment.

**Trading volume and Turnover rate.** The trading volume and turnover rate can reflect the liquidity of the market and the enthusiasm of the participants. Markets trade more frequently when investor sentiment is high, and vice versa.

**Number of IPOs and first-day yield.** In most cases, the higher the investor's attention to the stock, the higher the stock's first-day return. At the same time, investors' misunderstanding of public information interpretation or confidence in their own judgments may also lead to higher yields. Yi Zhigao and Mao Ning (2009) believed that the number of IPOs and the first-day returns of listing could better reflect the enthusiasm of investors, and they were both positive indicators of sentiment. [3]

**Number of new investor accounts.** The number of

new investors opening accounts represents how actively people participate in the stock market. Zhan Yubo (2002) found that OTC investors in China's stock market, especially small and medium investors, often lacked analysis of the economy and the fundamentals of listed companies, and only judged whether they should open an account based on the rise and fall of spot stock prices. [5] When investor sentiment is high, the willingness to enter the market is relatively higher. Therefore, the number of newly opened accounts can be used as a reference for index construction.

**Consumer Confidence Index.** Consumer Confidence Index is a comprehensive reflection and quantification of consumers' evaluation of the current economic situation and their subjective feelings about economic prospects, income levels, income expectations and consumer psychology. When people have positive expectations about the future economy and income, investment behavior is also more active, and vice versa. So Consumer Confidence Index can also reflect investor sentiment.

**Advance/Decline Line.** Advance/Decline Line is a cumulative value of the number of stocks that closed up each day minus the number of stocks that closed down. It is used to measure movement and trends in the stock market, but cannot describe changes in individual stocks. Under normal circumstances, when ADL and the stock market index rise at the same time, the upward trend of the market will continue, and investor sentiment will be relatively high at this time.

**Business Index of Macro-economic.** The Business Index of Macro-economic is a comprehensive statistical index reflecting the operation of the macro economy. It is also an indicator that reflects entrepreneurs' feelings and confidence in the macroeconomic environment and predicts the changing trend of economic development. Investors have more funds and better investment opportunities when the economic cycle is in an expansion phase. This is a high point for investor sentiment. In addition, the confidence of entrepreneurs is also the basis of market trust. Investors are obviously reluctant to invest funds in entrepreneurs with negative attitudes.

### 4. Index construction

Firstly, KMO and Bartlett's test are carried out to judge whether principal component analysis can be carried out.

**Table 1.** Kaiser-Meyer-Olkin and Bartlett's test (credit: original)

Kaiser-Meyer-Olkin	0.681
Bartlett's test	0.000***

\*\*\* represents the 1% significance level

Kaiser-meyer-olkin is an index used to compare simple correlation coefficients and partial correlation coefficients between variables, with values between 0 and 1. The closer the KMO value is to 1, the stronger the correlation between variables, and the more suitable the original variables are for factor analysis. The value of KMO of the test is 0.681.

At the same time, the results of the Bartlett's test showed that the significant P value was 0.000\*\*\*, which was significant at the level, rejecting the null hypothesis, and the variables were correlated. Analysis works.

**Table 2.** Explanation rate of the total variance (credit: original)

Elements	Characteristic root		
	Characteristic root	Percent variance	Accumulation
1	3.321	36.905	36.905
2	1.988	22.086	58.991
3	0.954	10.605	69.596
4	0.867	9.634	79.231
5	0.758	8.424	87.655
6	0.541	6.013	93.668
7	0.362	4.023	97.691
8	0.181	2.015	99.706
9	0.026	0.294	100

The above table is the total variance interpretation table, mainly to see the contribution rate of the principal components to the variable interpretation. In other words, it represents how many principal components are needed to express the variable as 100%. In general, the higher the variance explained rate, the more important the principal component is, and the higher the weight ratio should be.

In component 3, the characteristic root is lower than 1, and the contribution rate of variable explanation reaches 69.596%. However, the critical value of component 4 is relatively large, and the explanatory degree of 3 components is insufficient, so 4 components are selected.

**Table 3.** Factor Loading Table (credit: original)

	Factor loading				Communality
	F1	F2	F3	F4	
Closed-end Fund Discount Rate	0.649	0.132	-0.140	0.576	0.790
Trading volume	0.830	-0.470	-0.107	-0.149	0.943
Number of IPOs	0.578	0.507	0.130	-0.314	0.707
First-day yield of IPOs	0.459	0.377	0.291	-0.271	0.511
Number of new investors opened accounts	0.893	0.041	-0.060	-0.069	0.807
Business Index of Macro-economic Consumers	-0.044	0.796	0.063	-0.221	0.688
Confidence index Advance/Decline Line	0.402	0.620	0.037	0.487	0.785
Market turnover rate	0.839	-0.467	-0.052	-0.119	0.939

The above table is the factor loading coefficient table, which can analyze the importance of the latent variables in each principal component. F1, F2, F3, and F4 represent the newly constructed 4 components, respectively.

**Table 4.** Ingredient Matrix Table (credit: original)

Names	Ingredients			
	F1	F2	F3	F4
Closed-end Fund Discount Rate (a1)	0.195	0.066	-0.147	0.664
Trading volume (a2)	0.25	-0.236	-0.112	-0.171
Number of IPOs (a3)	0.174	0.255	0.136	-0.362
First-day yield of IPOs (a4)	0.138	0.19	0.305	-0.313
Number of new investors opened accounts (a5)	0.269	0.021	-0.063	-0.079
Business Index of Macro-economic (a6)	-0.013	0.4	0.066	-0.254
Consumers Confidence index (a7)	0.121	0.312	0.038	0.562
Advance/Decline Line (a8)	0.019	-0.169	0.943	0.218
Market turnover rate (a9)	0.253	-0.235	-0.054	-0.138

The a1 to a9 are used to represent different factors. Resulting in the following formula.

$$F1 = 0.195 \times a1 + 0.25 \times a2 + 0.174 \times a3 + 0.138 \times a4 + 0.269 \times a5 - 0.013 \times a6 + 0.121 \times a7 + 0.019 \times a8 + 0.253 \times a9$$

$$F2 = 0.066 \times a1 - 0.236 \times a2 + 0.255 \times a3 + 0.19 \times a4 + 0.021 \times a5 + 0.4 \times a6 + 0.312 \times a7 - 0.169 \times a8 - 0.235 \times a9$$

$$F3 = -0.147 \times a1 - 0.112 \times a2 + 0.136 \times a3 + 0.305 \times a4 - 0.063 \times a5 + 0.066 \times a6 + 0.038 \times a7 + 0.943 \times a8 - 0.054 \times a9$$

$$F4 = 0.664 \times a1 - 0.171 \times a2 - 0.362 \times a3 - 0.313 \times a4 - 0.079 \times a5 - 0.254 \times a6 + 0.562 \times a7 + 0.218 \times a8 - 0.138 \times a9$$

$$F = (0.369/0.792) \times F1 + (0.221/0.792) \times F2 + (0.106/0.792) \times F3 + (0.096/0.792) \times F4$$

In the following text, F will be used to represent the investor sentiment index.

## 5. Statistical analysis

### 5.1. Unit root test

**Table 5.** ADF Test (credit: original)

Variable	Differentia l order	t	P	AIC	Critical value		
					1%	5%	10%
F	0	-2.74	0.067*	568.46	-3.48	-2.88	-2.57
		5			1	3	8
	1	-7.68	0.000**	567.87	-3.48	-2.88	-2.57
		4			4	1	4
2	-6.76	0.000**	577.59	-3.48	-2.88	-2.57	
	5			2	4	5	9

\*\*\*, \*\*, and \* represent the significance levels of 1%, 5%, and 10%, respectively

The letter t represents the T-test result. And 'AIC' stands for Akaike information criterion. It is a standard for evaluating the complexity of statistical models and measuring Goodness of fit. The lower the number the better.

The results of the series test show that when the difference is 1 order, the significant P value is 0.000\*\*\*, which is significant at the level, rejecting the null hypothesis, and the series is a stationary time series. The same is true when the difference is of order 2, which is still a stationary time series.

A stationary time series needs to meet three conditions. First, the mean is finite and constant. Second, the variance is finite and constant. Third, the autocovariance is only related to the lag order. The significance of a stationary time series is that it can summarize laws and predict future possibilities through historical data.



**Figure 1.** Timing diagra (credit: original)

The above figure shows the timing diagram of the original data after 1st order difference.

## 5.2. Autoregressive Integrated Moving Average model

Autoregressive Integrated Moving Average model is one of the time series forecasting analysis methods. The ARIMA model requires that the residuals of the model have no autocorrelation and are white noise.

**Table 6.** ARIMA model (1,1,1) (credit: original)

	Symbols	Value
Sample size	N	137
Q statistic	Q6(P value)	0.018(0.894)
	Q12(P value)	2.202(0.900)
	Q18(P value)	6.531(0.887)
	Q24(P value)	13.294(0.774)
Information criterion	AIC	627.475
	BIC	639.126
Goodness of Fit	R <sup>2</sup>	0.684

\*\*\*, \*\*, and \* represent the significance levels of 1%, 5%, and 10%, respectively

BIC stands for Bayesian Information Criterion. It is similar to AIC, but the penalty term is larger than that of AIC, considering the sample size. When the number of samples is too large, it can effectively prevent the high model complexity caused by the high model accuracy.

From the analysis of the Q statistic results, it can be concluded that Q6 is not significant at the level, and the assumption that the residual of the model is a white noise sequence cannot be rejected. At the same time, the goodness of fit of the model R<sup>2</sup> is 0.684, the model performance is relatively good, and the model basically satisfies requirements.



**Figure 2.** The value graph (credit: original)

The blue line represents the true value, the green line represents the fitted value, and the yellow line represents the predicted value for the next seven months. It can be seen that the model can better track the true value and play a tracking effect.

**Table 7.** Model parameter table (credit: original)

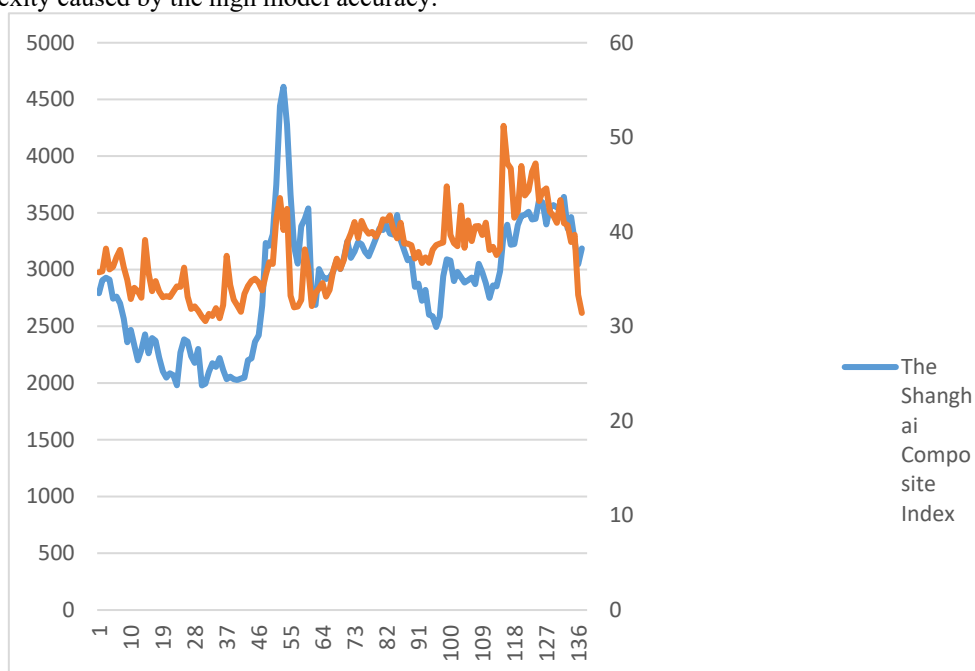
	Coefficient	Standard deviation	t	P> t
Constant	0	0.09	0.001	0.999
ar.L1.D.F	0.487	0.226	2.15	0.032
ma.L1.D.F	-0.785	0.182	-4.304	0

The model result is the ARIMA model (1,1,1) test table and is based on 1 difference data. The model formula is as follows.

$$y(t) = 0.487 * y(t - 1) - 0.785 * \epsilon(t - 1) \quad (1)$$

## 6. Relationship

### 6.1. Analysis of trend chart



**Figure 3.** Trend chart (credit: original)

As can be seen from the above figure, there is a certain relationship between the trend

of investor sentiment and the changes in market share prices. They both move in the same direction in most cases. In April 2015, February 2019, June 2020, investor sentiment reached several peaks. At the same time, the stock price continued to rise. However, when stock prices peaked, investor sentiment rose with it, but to a lesser extent. And investor sentiment is at a 10-year high in a subsequent rally in the stock. After 2022, investor sentiment is lower and market expectations are relatively pessimistic.

## 6.2. Spearman's rank correlation coefficient

**Table 8.** Spearman's rank correlation coefficient (credit: original)

	F	SSEC
F	1.000	0.735
SSEC	0.735	1.000

The Spearman correlation coefficient is a nonparametric measure of the dependence of two variables, and it does not assume that the two datasets are equally distributed. The correlation coefficient between investor sentiment and the Shanghai Composite Index is 0.735, indicating a comparatively strong correlation.

## 7. Conclusion

Based on the research by Baker M and Wurgler J and the construction methods of Chinese local researchers ISI and CICI, this paper selected 9 individual indicators and constructed a comprehensive index to measure the sentiment of Chinese investors. In order to be closer to the real reflection of investors, this paper does not eliminate the impact of the economic cycle and take the macroeconomy factor into consideration. Through the ADF test, this paper finds that the investor sentiment index meets the requirements of a stationary time series. Therefore, an ARIMA model (1,1,1) is established, which summarizes the laws of the index and tries to make predictions. Through the preliminary test, it is found that the index has a relatively strong relationship with the stock market.

This article has two shortcomings. First of all, the constructed index is insufficient to explain the price changes in the stock market, and cannot simulate the interaction between investor sentiment and stock price. Second, the Chinese stock market is immature, with many retail investors and high volatility. Data with shorter intervals is needed to construct indices, such as daily turnover or Advance/Decline Line. However, there are also difficulties in the acquisition and statistics of short-term data.

For investor sentiment, there are two alternative directions for the future. First, consider investor sentiment as a part of valuation, and no longer stick to rational person assumptions. On this basis, the multi-factor model or other asset pricing model can be revised, so that the

model has a higher degree of explanation for the real world. Second, from the perspective of behavioral finance, use data to explain the irrational behavior of investors.

## References

1. Baker, Malcolm, Jeffrey Wurgler, 2006, Investor Sentiment and the Cross-Section of Stock Returns, *Journal of Finance*61: 1645-1680
2. Klein, A., Riekert, M., Kirilov, L., Leukel, J. (2018). Increasing the Explanatory Power of Investor Sentiment Analysis for Commodities in Online Media. In: Abramowicz, W., Paschke, A. (eds) *Business Information Systems. BIS 2018. Lecture Notes in Business Information Processing*, vol 320. Springer, Cham. [https://doi.org/10.1007/978-3-319-93931-5\\_23](https://doi.org/10.1007/978-3-319-93931-5_23)
3. Yi Zhigao, Mao Ning. Research on investor sentiment measurement in China's stock market: the construction of CICI [J]. *Financial Research*, 2009 (11): 174-184
4. Wei Xingji, Xia Weili, Sun Tongtong. Research on Measuring Investor Sentiment in A-share Market Based on BW Model [J]. *Management Observation*, 2014 (33): 71-73+76
5. Zhan Yubo. Analysis on the Correlation between Chinese Stock Market Investors' Account Opening and Stock Market Fluctuation [J]. *Journal of Chongqing University of Business*, 2002 (04): 41-43
6. Wang Chunli, Liu Guang, Wang Qi. Multi factor quantitative stock selection model and timing strategy [J]. *Journal of Northeast University of Finance and Economics*, 2018 (05): 81-87. DOI: 10.19653/j.cnki.dbcjdxsb.2018.05.011
7. Wu Yanran, Han Liyan. Incomplete Rationality, Investor Sentiment and the Mystery of Closed end Funds [J]. *Economic Research*, 2007 (03): 117-129
8. Lan, D., Wu, S., Wei, G. (2022). The Influence of Investor Sentiment on Stock Market Based on Sentiment Analysis. In: Shi, X., Bohács, G., Ma, Y., Gong, D., Shang, X. (eds) *LISS 2021. Lecture Notes in Operations Research*. Springer, Singapore. [https://doi.org/10.1007/978-981-16-8656-6\\_31](https://doi.org/10.1007/978-981-16-8656-6_31)
9. Mathieu, A. (2016). The Impact of Investor Sentiment on REIT Returns. In: *Essays on the Impact of Sentiment on Real Estate Investments. Essays in Real Estate Research*. Springer Gabler, Wiesbaden. [https://doi.org/10.1007/978-3-658-11637-8\\_2](https://doi.org/10.1007/978-3-658-11637-8_2)
10. Liao, MH., Chen, YJ., Yu, CW., Chan, YL. (2022). Comparing Investor Sentiment Between Growth and Value Stocks. In: Barolli, L. (eds) *Innovative Mobile and Internet Services in Ubiquitous Computing. IMIS 2022. Lecture Notes in Networks and Systems*, vol 496. Springer, Cham. [https://doi.org/10.1007/978-3-031-08819-3\\_38](https://doi.org/10.1007/978-3-031-08819-3_38)
11. Weng, X., Luo, Y., Gao, J., Feng, H., Huang, K. (2019). The Comment of BBS: How Investor

Sentiment Affects a Share Market of China. In:  
Renault, É., Mühlethaler, P., Boumerdassi, S. (eds)  
Machine Learning for Networking. MLN 2018.  
Lecture Notes in Computer Science (), vol 11407.  
Springer, Cham.  
[https://doi.org/10.1007/978-3-030-19945-6\\_19](https://doi.org/10.1007/978-3-030-19945-6_19)