International Journal of Global Economics and Management
ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 5, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijgem.v5n3.19
Journal homepage: https://wepub.org/index.php/IJGEM/index

Analysis of Investor Sentiment and Trading Behavior

Junyu Long

School of Mathematics and Statistics, Central South University, Changsha, 410083, China
mathlongjy@126.com

ABSTRACT

This work focuses on analyzing the non-ferrous metal stock market by selecting specific indicators to reflect the sector's overall performance, considering the prevalent market conditions. In this paper, we preprocessed the index data and applied principal component analysis to obtain five principal components, accounting for 95.263% of the total variance, thereby establishing an investor sentiment measurement model. Further, to establish a correlation between trading volume and investor sentiment, we conducted logistic regression analysis based on the sentiment measurement model. The tested model achieved an accuracy rate of 87.3%. Additionally, using a Python web crawler, we collected and standardized the "greed and fear index" as an emotional indicator, matched it with daily data, and incorporated it into the logistic regression model. This enhanced model exhibited an accuracy rate of 82.5%. Our findings reveal that the greed and fear index has a more significant impact on emotional index changes than trading volume and turnover rate in logistic regression. This work not only contributes to a deeper understanding of investor sentiment in the non-ferrous metal stock market but also provides insights into potential market behaviors, guiding investors and policymakers in making informed decisions.

KEYWORDS

Stock prediction; Investor sentiment; Principal component analysis; Logistic regression

1. INTRODUCTION

Noise investors are irrational investors, and the transaction of irrational behavior is closely related to the current stock exchange market, which has become an important part of behavioral finance. Investor sentiment is an important factor that leads to irrational trading. The main body of investor sentiment includes both rational investors and irrational investors. In order to prove the existence of investor sentiment and measure the impact of investor sentiment on China's securities market, it is necessary to study the current situation of China's stock market, select the corresponding indicators to build the sentiment index, and find the correlation and asymmetric impact between investor sentiment and trading behavior.

In the stock market, investor sentiment and emotion play an important role in the volatility and trend of the market. There are two main types of emotions that play a dominant role, namely fear and greed. Fear is a common emotion for investors when faced with risk and uncertainty. Investors may exhibit fear when the market is faced with uncertainty, weak economic data, geopolitical tensions, or other negative events. This can cause investors to sell stocks and the stock market to fall. Greed is the emotion that investors feel when they seek high returns and a rally. When the market is performing strongly, market sentiment is optimistic, or investors see stocks that are running high, greed can trigger investor speculation and excessive optimism. This can lead to rising stock prices and bubbles in the market. In the traditional stock market, the Fear & Greed index is a way of gauging where the stock market is going and whether stocks are reasonably priced. Excessive fear tends to depress stock prices, while excessive greed tends to have the opposite effect.

2. SOURCES AND METHODS OF DATA

2.1. Data Sources

The stock trading data of some listed companies in the non-ferrous metals sector of GF Securities from January 4, 2021 to March 31, 2023 are selected as the research data set.

2.2. Theoretical Basis

(1) KMO test and Bartlett test

The KMO measure measures the correlation between variables by calculating the ratio of the partial variance of the observation to the total variance of the observation.

$$\text{KMO} = (\sum_i (P_i)^2) \, / \, (\sum_i (P_i)^2 + \sum_i (P_i')^2) \tag{1}$$

Where, is the Pearson correlation coefficient between the variables; $P_i P_i'$ Is the partial correlation coefficient, that is, the correlation coefficient calculated after controlling the influence of other variables.

The KMO value ranges from 0 to 1, and close to 1 indicates that the variables have a good relationship with each other and are suitable for principal component analysis; While close to 0 indicates that the correlation between variables is weak and is not suitable for principal component analysis. In general, a KMO value greater than 0.6 is considered an acceptable threshold.

The Bartlett test is based on the null hypothesis (H0) and the alternative hypothesis (H1) of the covariance matrix of the observed data.

H0: The covariance matrix of the observed data is an identity matrix, that is, there is no correlation between the variables.

H1: There is at least one correlation.

The statistics of the Bartlett test are calculated as follows:

$$X^2 = -2 \times (\text{F-} (\sum_i \ln |P_i|)) \tag{2}$$

Where F is the degree of freedom of the correlation matrix, and || represents the absolute value of the correlation coefficient P_i.

According to the calculated Bartlett test statistics, the critical values of corresponding degrees of freedom and significance level can be found by referring to the Chi-square distribution table, so as to judge the significance of Bartlett test.

(2) Principal component analysis

Principal Components Analysis (PCA) is a dimensionality reduction algorithm, it can convert multiple indicators into a few principal components, these principal components are linear combinations of the original variables, and are unrelated to each other, which can reflect the majority of the original data information. In general, when the research problem involves multiple variables and there is a strong correlation between the variables, consider using the method of principal components analysis to simplify the data.

(3) Logistic regression

Logistic Regression is a statistical method used to build a classification model. It converts a linear combination of input variables to the probability of predicting a binary output based on a logical function (also known as a Sigmoid function). Logistic regression is often used to deal with binary classification problems, but can also be used to deal with multiclass problems by modifying a multiclass model.

3. THE ESTABLISHMENT OF INVESTOR SENTIMENT MEASUREMENT MODEL

3.1. Index Selection and Data Preprocessing

Through the selection of indicators, it reflects the difference between the individual stocks of non-ferrous sectors and the whole. Combined with literature and data[1, 2], daily capital return rate, average daily market return rate of equal weights, weighted average daily market return rate of total market value, weighted average daily capital return of total market value, weighted average daily capital return of circulating market value, daily risk-free rate of return, price/earnings ratio, price-to-sales ratio, price-to-book ratio and price-to-cash ratio are selected as indicators to reflect the whole non-ferrous sector.

Perform data preprocessing on the selected index data:

(1) For the missing data in the data, due to the large amount of overall data, it is chosen to abandon, which will not have much impact on the overall data.

(2) For outliers, the principle of interquartile range (IQR) is used for processing.

(3) In order to reduce the influence between the number of stocks of different sizes, we standardized the data with Z-Score.

3.2. Construct an Investor Sentiment Measurement Model

After analysis, 5 principal components can well explain all independent variables, can represent 95.263% of the whole independent variables. Assuming daily return on capital, equal weight average daily market return, price/earnings ratio, weighted average daily market return of total market value, price-to-sales ratio, price-to-book ratio, daily risk-free return, weighted average daily capital return of total market value and weighted average daily capital return of circulating market value are X1, X2, X3, X4, X5, X6, X7, X8, X9 and X10, respectively. The final results are as follows:

$$F1=0.125*X1+0.228*X2+0.005*X3+0.249*X4-0.002*X5-0.001*X6-0.002*X7+0.001*X8+0.249*X9+0.248*X10$$

$$F2=0.001*X1+0.0005*X2+0.17*X3-0.001*X4+0.324*X5+0.317*X6+0.324*X7+0.026*X8-0.001*X9-0.00006*X10$$

$$F3=0.007*X1-0.009X2+0.344*X3-0.004*X4-0.087*X5-0.0536*X6-0.115*X7+0.909*X8-0.001*X9-0.0005*X10$$

$$F4=1.057*X1-0.034*X2+0.14*X3-0.167*X4-0.032*X5-0.02*X6-0.0368*X7-0.072*X8-0.167*X9-0.168*X1$$

$$F5=0.164*X1+0.002*X2-0.991*X3-0.022*X4+0.155*X5+0.139*X6+0.189*X7+0.421*X8-0.021*X9-0.020*X10$$

By extracting the sum of squares and loading coefficients from the explained total variance table, the relationship between the comprehensive principal component F and the above 5 principal components F1, F2, F3, F4 and F5 can be obtained, namely:

$$F=(0.395/0.953)*F1+(0.294/0.953)*F2+(0.103/0.953)*F3+(0.081/0.953)*F4+(0.079/0.953)*F5$$

3.3. Model Evaluation

KMO and Bartlett's tests are performed to see if the sequence data is suitable for principal component analysis. The results of KMO test show that the value of KMO is 0.804; meanwhile, the results of Bartlett sphericity test show that the significance P value is 0.000***, showing significance at the level, rejecting the null hypothesis, correlation among variables, and the principal component analysis is valid and the degree is suitable [3, 4].

4. THE ESTABLISHMENT OF THE CORRELATION MODEL OF INVESTOR TRADING BEHAVIOR, DATE AND INVESTOR SENTIMENT

4.1. Analysis of Investor Sentiment

According to the established sentiment measurement model, the investor sentiment analysis is carried out for the whole non-ferrous stocks, and finally three classification indicators are obtained.

Table 1. Basic summary of multiple classification dependent variables

Dependent Variables	Options	Frequency	Percentage (%)
emotion	Rocking	447	82.472
	Optimism	53	9.779
	Pessimism	42	7.749
	Totals	542	100

4.2. Model the Correlation of Investor Trading Behavior, Date and Investor Sentiment

According to the above classification indicators, two binary logistic regression models are trained, each model corresponds to a category, and the probability of each category is calculated and the classification decision is made when forecasting.

(1) Analysis Based on Pessimism vs. Indecisiveness

Constant Effect: The constant term exhibits statistical significance at the 0.000*** level ($P < 0.001$), rejecting the null hypothesis. This indicates that the constant has a notable impact on sentiment, translating into a 1048.63% increase in the likelihood of indecisiveness over pessimism for each unit increase in the constant.

Date Effect: The date variable also demonstrates significance at the 0.025** level ($P < 0.05$), suggesting a significant influence on sentiment. Specifically, as the date progresses by one unit, the probability of indecisiveness decreases by 33.274% relative to pessimism.

Daily Average Trading Volume: The daily average trading volume fails to reach statistical significance ($P = 0.961$), thereby failing to reject the null hypothesis. Consequently, this variable does not significantly affect sentiment.

Daily Average Turnover Rate: Similarly, the daily average turnover rate does not exhibit statistical significance ($P = 0.449$), indicating no notable impact on sentiment.

(2) Analysis Based on Pessimism vs. Optimism

Constant Effect: In contrast to the previous framework, the constant term does not show statistical significance ($P = 0.454$), suggesting no significant influence on sentiment when comparing pessimism to optimism.

Date Effect: Notably, the date variable retains its statistical significance at the 0.000*** level ($P < 0.001$), indicating a profound effect on sentiment. Specifically, as the date advances by one unit, the probability of optimism decreases by 59.629% compared to pessimism.

Daily Average Trading Volume: Consistent with the previous analysis, the daily average trading volume remains statistically insignificant ($P = 0.185$), indicating no discernible impact on sentiment dynamics.

Daily Average Turnover Rate: The daily average turnover rate, albeit approaching significance at the 0.10 level (P = 0.080*), does not achieve statistical significance, thereby not influencing sentiment in a meaningful manner [5, 6].

4.3. Model Evaluation

The results of the Chi-square test of the likelihood ratio of the model show that the significance P value is 0.001***, showing significance on the level, rejecting the null hypothesis, so the model is valid. The AUC value of the classification evaluation index is 0.923, indicating that the classification effect is good, and the accuracy rate is as high as 87.3%, it can be seen that the effect of the logistic regression is good [3, 7, 8].

5. THE IMPACT OF FEAR AND GREED INDEX ON INVESTOR SENTIMENT

5.1. Data Collection and Processing

By crawling through the US-CNN Fear & Greed Index data, the data is usually provided by third-party institutions or financial websites to measure investor sentiment and the emotional state of the market, with authoritative and reference value. (https://sc.macromicro.me/charts/50108/cnn-fear-and-greed)

On the basis of using python web crawler to crawl the Fear Greed Index taken from 2021-01-01, the obtained Fear Greed Index is graded:

0-20: Extreme Fear - indicates that investor sentiment is very fearful and pessimistic. The market may be considered over-sold and investor sentiment may be very subdued.

20-40: Fear - indicates that investor sentiment is still relatively fearful, but not as much as extreme fear. The market may be in a correction or downtrend.

40-60: Neutral - indicates that investor sentiment is relatively balanced, with no overt fear or greed dominating. The market may be in a sideways or uncertain phase.

60-80: Greed - indicates that investor sentiment is trending toward greed and optimism. The market may be considered overbought and investors may be overly chasing the rally.

80-100: Extreme Greed - indicates that investor sentiment is extremely greedy and optimistic. The market may be considered an overbought and highly risky state.

Add the fear-greed index and rank results to the original dataset by matching them with dates. Outliers and missing values in the data were processed, and Z-Score normalization was performed.

5.2. Establish a Correlation Model Between Investor Trading Behavior, Date, Fear And Greed Index and Investor Sentiment

Similarly, two binary logistic regression models are trained according to the classification indexes.

(1) Analysis Based on Pessimism vs. Indecisiveness

Constant Effect: The constant term exhibits a highly significant P-value of 0.000***, indicating statistical significance. This rejects the null hypothesis and suggests that the constant has a pronounced impact on sentiment. Specifically, for each unit increase in the constant, the probability of indecisiveness surges by 1186.059% relative to pessimism.

Date Effect: The date variable displays a borderline significance with a P-value of 0.070*, failing to achieve statistical significance at the conventional levels. Thus, the null hypothesis cannot be rejected, and it is concluded that the date does not significantly influence sentiment.

Daily Average Trading Volume: With a P-value of 0.607, the daily average trading volume does not exhibit statistical significance, leading to the inability to reject the null hypothesis. Consequently, this variable does not significantly affect sentiment.

Daily Average Turnover Rate: Similarly, the daily average turnover rate fails to reach statistical significance (P = 0.913), indicating no notable impact on sentiment.

Greed and Fear Index (GFI) Effect: The GFI demonstrates statistical significance at the 0.002*** level, rejecting the null hypothesis. This suggests that the GFI has a significant influence on sentiment, with each unit increase in the GFI resulting in a 79.22% increase in the probability of indecisiveness over pessimism.

(2) Analysis Based on Pessimism vs. Optimism

Constant Effect: The constant term does not show statistical significance (P = 0.286), failing to reject the null hypothesis. Therefore, the constant does not significantly affect sentiment when comparing pessimism to optimism.

Date Effect: The date variable exhibits statistical significance at the 0.001*** level, rejecting the null hypothesis. This indicates that the date has a notable impact on sentiment, with each unit increase in the date resulting in a 56.749% decrease in the probability of optimism relative to pessimism.

Daily Average Trading Volume: Consistent with previous findings, the daily average trading volume remains statistically insignificant (P = 0.535), suggesting no discernible effect on sentiment.

Daily Average Turnover Rate: The daily average turnover rate also fails to reach statistical significance (P = 0.454), indicating no significant impact on sentiment dynamics.

Greed and Fear Index (GFI) Effect: The GFI displays statistical significance at the 0.004*** level, rejecting the null hypothesis. This underscores the GFI's significant influence on sentiment, with each unit increase in the GFI leading to a 100.951% increase in the probability of optimism over pessimism.

Therefore, the influence of the Greed and fear index on the emotion index will be greater than the influence of the daily average trading volume and the daily average turnover rate on the emotion index. Therefore, when analyzing the emotion index, the greed and fear index can be added for analysis.

5.3. Model Evaluation

The results of the likelihood ratio Chi-square test of the model show that the significance P value is 0.000***, showing significance at the level, rejecting the null hypothesis, so the model is valid. The AUC value is 0.907 and the accuracy rate is 82.5%, indicating that the relationship between the dependent variable and the independent variable can be explained well [3, 9, 10].

6. RESEARCH CONCLUSIONS AND SUGGESTIONS

6.1. Research Conclusion

By selecting the stock trading data of some listed companies in the non-ferrous metals sector of GF Securities from January 4, 2021 to March 31, 2023, combining principal component analysis and logistic regression model, this study deeply discusses the impact of investor sentiment on the stock market of non-ferrous metals sector, and establishes the measurement model of investor sentiment and its correlation model with trading behavior. The main conclusions are as follows:

(1) Construction of investor sentiment measurement model:

Through principal component analysis, five principal components (F1 to F5) were extracted from multiple financial indicators. These principal components could represent 95.263% of the overall independent variable information, effectively reducing the data dimension and retaining key information.

The investor sentiment measurement model is constructed, which integrates multiple financial indicators such as daily capital return rate, average daily market return rate of equal weights, price-earnings ratio, price-sales ratio, price-to-book ratio, etc., and can reflect the overall emotional state of the non-ferrous metal sector in a more comprehensive way.

(2) The correlation between investor sentiment and trading behavior:

Through logistic regression analysis, the correlation model between investor sentiment and trading behavior (including average daily trading volume and average daily turnover rate) is established. The results show that although the effect of trading volume and turnover rate on sentiment is not significant, the date has a significant effect on sentiment, indicating that market sentiment fluctuates with time.

After further introducing the Greed and fear index, it is found that the index can significantly affect the change of sentiment index compared with the trading volume and turnover rate, which verifies the important role of greed and fear index in measuring market sentiment.

(3) Model validation and effect evaluation:

The accuracy of the initial model reached 87.3%, indicating that the established model of investor sentiment and its correlation with trading behavior has high prediction accuracy.

After the introduction of the greedy-fear index, the accuracy of the model decreased slightly but remained at 82.5%, indicating the effectiveness of greedy-fear index as an emotional indicator and enhancing the explanatory ability of the model.

6.2. Suggestions

Based on the above research conclusions, the following suggestions are made:

(1) Strengthen investor sentiment monitoring:

Financial institutions and regulators should strengthen the monitoring and analysis of investor sentiment, use principal component analysis, logistic regression and other statistical methods to build sentiment index models, timely grasp the changes in market sentiment, and provide scientific basis for policy formulation and market supervision.

(2) Improve investor education:

Strengthen investor education to improve investors' risk awareness and rational investment ability. Through publicity and education activities, investors should be guided to establish correct investment concepts and reduce the occurrence of irrational trading behaviors.

(3) Pay attention to changes in market sentiment:

Investors should pay attention to changes in market sentiment when making investment decisions. When market sentiment is too optimistic or pessimistic, they should remain vigilant and avoid blindly following the trend or panic selling. At the same time, sentiment indicators such as the Greed and Fear index can be used as auxiliary tools to make investment decisions more scientific and accurate.

(4) Improve market information disclosure mechanism:

Improve the market information disclosure mechanism and enhance market transparency. Market information should be disclosed in a timely, accurate and comprehensive manner to reduce the occurrence of asymmetric information and reduce the impact of market noise on investor sentiment.

(5) Strengthen international cooperation and exchanges:

Strengthen cooperation and exchanges with the international financial market, learn from international advanced experience and technical means, and enhance China's ability to monitor and analyze financial market sentiment. At the same time, we will actively participate in the formulation and improvement of international financial market rules to create a more favorable external environment for the development of China's financial market.

REFERENCES

[1] Yi, H., Cai, Y., & Dong, D. (2017). Research on construction and influence of investor sentiment index in Chinese stock market. Price Theory and Practice, (10), 4.

[2] Du, X., & Li, Y. (2016). A study on the construction of stock investor sentiment index in China. Economic and Trade Practice, (2X), 2.

[3] Gao, H. (2005) Application of multivariate statistical Analysis. Peking University Press, Beijing.

[4] Bandopadhyaya, A., & Jones, A. L. (2006). Measuring investor sentiment in equity markets. Journal of Asset Management, (7), 208-215.

[5] Almansour, B. (2015). The impact of market sentiment index on stock returns: An empirical investigation on Kuala Lumpur Stock Exchange. Journal of Arts, Science & Commerce, 6(3), 1-28.

[6] Hu, J., Sui, Y., & Ma, F. (2021). The measurement method of investor sentiment and its relationship with stock market. Computational Intelligence and Neuroscience, 1-11.

[7] Huang, D., Wen, F., & Yang, X. (2009). Investor sentiment index and empirical analysis of Chinese stock market. Systems Science and Mathematics, (1), 13.

[8] Liang, C., Tang, L., Li, Y., & Wei, Y. (2020). Which sentiment index is more informative to forecast stock market volatility? Evidence from China. International Review of Financial Analysis, (71), 101552.

[9] Yuan, X., Zhang, P., Xiong, S., et al. (2021) Research on Identification Method of abnormal Power Consumption Based on Logistic Regression Algorithm. Electrical Measurement and Instrumentation, 58(12):81-87.

[10] Yoshinaga, C. E., & Castro Junior, F. H. F. D. (2012). The relationship between market sentiment index and stock rates of return: A panel data analysis. BAR-Brazilian Administration Review, (9), 189-210.