Research on Influence from Investor Focus and Investor Sentiment in On-line Social Media upon Stock Market

Yingchen Lu

Nanjing University Business School, Nanjing 210000, Jiangsu, China

Abstract: The present investor sentiment detection relies mainly on the data from security exchange market, such as the discount of closed-end fund, the premium rate on the first day of IPO and other indexes. Although they can directly substitute for investor sentiment, those indexes lack objectivity and timeliness, and thus cannot provide precise evaluation over investor sentiment. With the development of information technology, people gradually engage in recreational activities and make friends via the Internet, and express their own opinions based on the on-line information. In this regard, the influence from investor focus and investor sentiment in the on-line social media upon stock market has been explored from the perspective of big data on line. According to analysis on investor sentiment and dynamics concerning stock market earnings, those two indexes are positively correlated and the stock market fluctuates sharply in a short period of time. Sound countermeasures should be taken to enhance the safeguard against such fluctuations.

Keywords: on-line social media, investor focus, investor sentiment, stock market and influence

1. Introduction

Stock market is an emerging market in China during recent years. With the construction of stock market, the blind individual investment and the sell into corrections has grown to the most prominent problems. The factors influencing stock market performance have gradually become the main concern among stock investors. However, stocks are featured by their high risk and high yield. As on-line platforms arise, people begin to pay attention to stock market dynamics and carry out transactions on those on-line social platforms. In light of those problems, various new network techniques have been adopted to analyze and develop investor sentiment indexes based on big data market, and references have been provided for optimizing securities market environment and enhancing management through measurement of the relationship between investor sentiment index and stock market.

2. Overview on investor sentiment

There are some scholars researching investor sentiment from the perspective of behavioral finance. They judge the influence from investor sentiment index upon the price fluctuation in the securities market through measuring those indexes. The president of American Academy of Financial Management once expounded the noise trader by introducing investor sentiment into the securities market. He proposed that investor sentiments are inter-playing to guarantee a stable yield for both. However, it's impossible to evaluate the fault orientation caused by irrational behavior during this process. In this regard, investor sentiment will become a systemic risk impacting the stability of financial assets. At present, the main methods to measure investor sentiment are indirect index and direct index. Indirect index works as a variable factor to evaluate investor sentiment through financial market. In addition, American scholars also proposed representative indexes of discount of closed-end fund, zero-stock transaction rate and mutual fund redemption rate. According to the analysis on investor sentiment through this index, the final yield rate of stock market can be effectively expected. Generally speaking, for stocks with a high proportion of individual ownership, the investor sentiment index needs to be created upon transaction data provided by stock exchange. It turns out to be that the lower the proportion of penny stocks, small-cap stocks and institutional investors is, the easier for those stocks to obtain high yield, which has less influence on investor sentiment. Indirect index mainly refer to financial market information featured by its simple data collection and strong objectivity. However, indirect index can only works as a representative index of investor sentiment so there are certain errors. Generally, direct index can be obtained through social survey that mainly investigates investment information issued in the market or by governmental agencies [1-2].

3. Stock data sources and research methods

Investor sentiment has been analyzed and summarized mainly from the stock evaluation information in the social media and the polling data on line. Currently, each major portal website, and financial and economic website in China have opened their stock forums and provided various discussion and communication approaches for stock investors. A financial and economic website has been selected according to the analysis on the page view and post volume of daily investors, time and other factors. Social media data sources have been analyzed to extract investor sentiment index. Crawlers can be adopted to analyze the comments, posts, clicks and other information concerning stock investors so as to analyze the data effectively. The above information can truly reflect the investor sentiment. In addition, all the financial and economic websites also regularly publish the information on stock market and guide netizens to vote so as to accurately reflect investors' opinions on market trends in the future [3].

3.1 Analysis on Text Sentiment Tendency

With the development of network technologies, natural language analysis method has been gradually adopted by multiple industries. It's very reliable to analyze the characteristics of investors by this method. The specific ways are mainly as follows in general: I. Create an investor sentiment management library to manage, summarize and analyze the ideas and suggestions posted by investors on line, so as to draw final evaluation conclusions. II. Adopt computer equipment to calculate and learn algorithms so as to complete the sentiment library. This technology is mainly applied into business evaluation and can generate final results effectively, which is better than sentiment word dictionary method. III. Affective polarity categorization by sentiment word dictionary. This method analyzes the information on positive and negative emotions in the text based on the compiled sentiment polarity dictionary so as to evaluate the sentiment orientation of document. It's easy to operate and thus can be widely applied. However, the method still has the following problems: There has been no complete sentiment word dictionary in China because the text analysis starts late. The NTUSD dictionary is more popular at present. However, this dictionary still cannot be adopted to analyze investor sentiment due to the immature financial market in China. That is because there are huge differences between Chinese text and English text, among words to describe emotions, as well as among specific situations where investor sentiment orientation is evaluated.

3.2 Emotional Judgment on Stock Information

When it comes to the specific analysis on information from investor message, the emotional tendencies of stock comments by investors should be judged first. Generally speaking, stock market traders will process all information before opening quotation, directly eliminate the information not posted by investors, restructure the comments by investors, delete some individualistic words, then normalize the text sentiment and finally analyze the investor sentiment. The specific steps are shown as follows: I. Data pre-processing. The information on comments in stock forums is sorted out through crawler. The information on investor sentiment is collected according to the practice during market opening. The final conclusions are drawn after the irrelevant posts and marketing information are eliminated. II. Text segmentation. Posts need to be processed through word segmentation in light of the difference between Chinese and English text information. Generally, a third-party package is selected for such processing so as to timely remove those obsolete words. III. Mark training data-set manually. Select some information from the comments on stocks for manual sentiment polarity marking and use the information as training data samples. During this process, in order to avoid human influence, some intermediaries need to be selected to mark the emotional tendency of the text and divide it into different emotions. Finally, the mode is selected as the result. IV. Feature representation. The text should be set to a recognizable form so as to recognize the information contained effectively. The main form adopted now is vector space model. Each word behind a participle in a text can become a spatial dimension. Weight algorithm has been adopted to analyze the status of words to text. V. Feature selection. According to the statistics, the median number of posts is large. The direct calculation will impact the categorization efficiency of the algorithm. Therefore, it's necessary to screen out those features valuable to sentiment information categorization before machine learning. Generally, the information gain method can be adopted for screening. Many words can be selected to reflect investor sentiment. VI. Machine learning algorithm is adopted to select experimental data; the experimental data are categorized through cross validation method. Finally, the results are obtained according to the classification algorithm. It can be found from the above analysis that the accuracy of machine algorithm is higher than that of sentiment word dictionary [4].

3.3 Construction of Investor Sentiment Indexes

In analyzing investor sentiment, one first will generally analyze the comments that have been processed through text sentiment marking. However, it should be noted during the specific processing that the market should not be opened on weekends in general, so as to analyze investor sentiment by virtue of information on the weekend, and to process those

emotions on Monday for subsequent utilization. There are two main ways for construction: First, CCTV watch index preparation method; the investor sentiment on the stock forum on line has been measured through bullish index. Second, the joint bull and bear survey method is adopted for processing. Those two aforesaid methods can accurately work out the time rules of investor sentiment. They are mainly two types of investor sentiment obtained from on-line social media and on-line voting [5].

4. Empirical research and analysis on relevant results

4.1 Model Analysis and Descriptive Statistics on Samples

Stock investor sentiment on line will change with transactions on the securities market. Investors will express their own opinions concerning development tendencies on the stock market at present and in the future. Those opinions will have a direct influence on their investment behaviors and decision-making views. There is a difference in the amount of time that investors spend watching stock market practices and making investment decisions. Therefore, the development tendency of stocks and the performance of on-line investors may also influence each other. In this regard, the vector regression model has been selected to calculate the relationship among the investor sentiment reflected in on-line social media and on-line voting, the ups and downs on the securities market, and the fluctuations of transaction volume, so as to construct a function model. Finally, a vector autoregressive model has been constructed in the time sequence. This model can accurately reflect the fluctuation extent in the Chinese securities market during recent years. According to the on-line investor sentiment index, the performance of investment sentiment is generally in alignment with the stock market tendency. The analysis on relationship shows that the investor sentiment in the website stock forum is positively correlated with the yield rate of Shanghai Composite Index. Investor sentiment in on-line voting is positively correlated with stock rise.

4.2 Stability Test and Hysteresis Selection in the Time Sequence

The time sequence of data should keep stable during model construction and result verification. Therefore, the investor sentiment index and the securities market variables should be tested. According the test results, the large variation difference between variables is mainly concentrated in a certain hypothetical interval. Therefore, the overall time sequence is found to be smooth and stable. In terms of hysteresis results of the model, it can be found that their indexes are all concentrated on a certain hysteresis stage. Hence, the model is stable and can be adopted for standard deviation function analysis.

4.3 Impulse Response Function

In a traditional causality, one variable can only be clarified by another. However, the interaction between the variables cannot be determined. Impulse response function is to exert impact force on a dynamic unit or organization so as to research its influence on the future changes in systematic variables. Thus, the impulse response function can also be adopted herein to judge the relationship between on-line investor sentiment and securities market. According to the research, when the investor sentiment index in the stock forum is positive, the comments from investors are more proactive. Meanwhile, transaction volume in the stock market will gradually increase. However, this phenomenon is only a short-run occurrence and will decline later. According to the investigation on bull and bear investor bullish index on the website, this fluctuation is constant and will eventually grow smooth. Thus, the investor sentiment on line has a positive influence on the yield of securities market index and it is mainly concentrated in a certain time period. In addition, according to the bull and bear survey to financial and economic websites and stock forums, when receiving a positive impact, the investor sentiment does not bring about a significant change in stock market turnover. However, a positive response will gradually occur in the later stage and then decrease after increasing to a certain height. In brief, the total response time of turnover is higher than the influence time of yield rate on the stock market [6].

4.4 Four-factor Model Created upon Investor Sentiment on Line

A three-factor model has been selected to analyze the dynamic tendency in Chinese stock market so that we can understand better the influence of investor sentiment on line upon stock market fluctuations. American researchers once researched the influencing factors of difference in stock portfolio return rate in the American stock market. The results show that the market value and the book value of listed companies exert the greatest influence. In this research, 55 investment portfolios have been selected for testing. The traditional three-factor model has first been adopted to simulate the yield rate of the portfolio comprising stocks. Then, investor sentiment variable has been added in to form a four-factor model. Finally, the influence from investor sentiment upon the yield rate of investment portfolio has been examined. The data of traditional three-factor model comes from abroad and are mainly calculated by econometric model expression. It has been found that the accuracy of simulation results from traditional three-factor model is up to 90% and can effectively reveal the yield rate of

investment portfolio. However, there is still some degree of deviation due to fluctuations in the stock market. The four-factor model formed by the introduction of investor sentiment variable shows a higher accuracy, which can effectively verify that investor sentiment has a positive influence on the yield rate of the Chinese stock market.

4.5 Final results

The information concerning on-line stock comments from the ordinary websites and the stock forums on financial and economic websites has been collected and investigated herein. Then, a complete system of investor sentiment index has been created through text analysis, machine learning algorithm and sentiment tendency. Model simulation method, causality analysis method, four-factor model and other methods have been adopted to analyze the relationship between the investor sentiment in the on-line social media and on-line voting and the yield rate of stock market, and the fluctuations and the turnover of Shanghai Composite Index. Thus, the following results have been obtained: (1) Machine learning algorithm is superior to sentiment word dictionary, and can categorize and analyze the sentiment in the text effectively. (2) Investor sentiment in the on-line social media has a direct influence on the turnover and the yield rate on the securities market. According to the research, when stock investors in stock forums are posting positive information, the yield rate of Shanghai Composite Index will rise, too. However, the Index will also fall quickly later. The bullish index in the bull and bear surveys from financial and economic website shows a similar fluctuation tendency as the above. Thus, it can be concluded that the analytic results on investor sentiment are relatively accurate in the short term, which further indicates that there is the sheep-flock effect in the Chinese stock market. Individual investor sentiment can also be subject to other investment comments in the on-line social media. Investors will spur yield rate to change through sell into corrections, which is featured by remarkable reversal and falling back. (3) Internet investor sentiment works as the main factor influencing the stock market turnover. If the influence lasts beyond 10 days, the influence on the stock market turnover is greater than that on yield. The subsequent transaction activity in the stock market will change upon investor sentiment. (4) The four-factor model has been adopted to analyze the on-line investor sentiment, which is more rational than the traditional three-factor model. Therefore, the influence from future changes in the Chinese stock market is small but comments from investors can also be adopted as a kind of reference to analyze the fluctuations on the stock market [7].

5. Summary

In brief, the research on the influence from on-line investor sentiment upon stock market can provide a basis for China improving its stock market system. With the improvement of quality in investors and netizens in China, there will be more valuable stock information from on-line stock comments in the on-line social media in the future. The text analysis method and the emotional tendency categorization method adopted herein has advanced the research on investor sentiment, and influenced the relationship between investor sentiment and stock market yield to a certain extent. In addition, in light of the continuous development of on-line social media, the comments by netizens in stock forum on financial and economic websites still need to be regulated and guided, so as to provide reference for the research on other stock markets.

References

[1] Shi Yong, Tang Jing & Guo Kun. Influence from Investor Focus and Investor Sentiment in Social Media on China's Stock Market[D]. University of Chinese Academy of Sciences, 2017.
[2] Qiu Jiangnan & Ge Yidi. Research on Influence from Investor Sentiment in Social Media on Stock Market under Stock Market Crisis[J]. *Management Review*. 2021, 33(5): 14.
[3] Zhang Jianjun. Research on Influence from Investor Sentiment on Stock Market under Background of Internet Finance[J]. *Finance and Economics*. 2020, (16): 4.
[4] Chen Bimei. Research on Influence from Investor Sentiment on Closing Price on the Stock Market[J]. *Journal of Jilin Radio and TV University*. 2020, (5): 2.
[5] Yang Huicui. Research on Influence from Chinese Investor Sentiment on A Share Market[J]. *Sci-Tech & Development of Enterprise*. 2019, (1): 184-185.
[6] Shi Yong, Tang Jing & Guo Kun. Influence from Investor Focus and Investor Sentiment in Social Media on China's Stock Market[J]. *Journal of Central University of Finance and Economics*. 2017, (7): 9.
[7] Xu Tianyang. Research on Influence from Investor Sentiment in On-line Social Media on Stock Market[J]. *Shanghai Management Science*. 2018, 40(3): 8.