Analytical Study on Fluctuations of Investor Network Public Opinion Information Evolvement in China
Panpan Yang\textsuperscript{a,*} and Xiaotong Li\textsuperscript{b}
Xi’an Technological University, Xi’an, Shanxi 710021, China
\textsuperscript{a}1571832291@qq.com, \textsuperscript{b}kelin7777@126.com
*Corresponding author

Keywords: Network Public Opinion, Investor, Fluctuation.

Abstract. Fluctuations of investor network public opinion information evolvement were adopted as the research object in this paper. By network information capture, text classification and sentiment scoring, etc., a comprehensive index was set for network public opinion fluctuations; furthermore, an EGARCH model was utilized to perform public opinion information fluctuation fitting. As indicated by fitting results, fluctuations of investor network public opinions exhibited some characteristics of volatility clustering and asymmetry, etc. so that they were more susceptible to impacts of negative news.

1. Introduction
Evolution of network public opinion information is a complicated dynamic process. Through researches on dissemination, evolvement rules and inherent mechanism of network public opinion information, not only are the formation law and the evolvement process of network public opinions grasped, but public opinion tendencies reflected on the Internet can be monitored, estimated and warned in advance, which is of great theoretical significances and practical values to effectively supervise and control network public opinions. At present, an increasing number of investors in China gather together on internet forums to discuss stock market information; as a consequence, drastic volatility of asset prices incurred by network public opinion fluctuations thus formed takes place occasionally. On this basis, fluctuation evolvement features of investor network public opinions in China were used as the research object to establish the corresponding model and determine parameters and properties of such a model by obtaining network public opinion information so as to further comprehend different states of public opinions in the process of evolvement. Without doubt, this is beneficial for the government to know about stability of investors and further take proactive measures to avoid the occurrence of unexpected group events that may threaten financial system safety.

2. Network Public Opinion Index Construction for Investor Sentiment
GARCH model was utilized in this paper to explore features and mechanism of public opinion evolvement based on the obtainment of investor network public opinions in China. Before the index has been constructed for investor network public opinions in China, an appropriate network platform was selected in the first place. As a socialized media that is dominated by text information and has the most intensive investor interactions in China, Guba is currently featured with being crowed interactively, a great amount of information and quick updating, etc. Therefore, platform Guba was selected in this study to perform data analysis mining. To be specific, index construction for investor network public opinions in China should follow several procedures below.
3. Information Acquisition

At present, there is plenty of crawler software that can be utilized for network information mining. In this paper, R Programming Language served as a crawler tool and R is a kind of very practical software. As open source software, it contains thousands of R packages provided by developers in all industries for free. Therefore, R software is provided with many powerful functions and diverse open source software packages are updated unceasingly. Two web crawlers selected for use from R in this paper were RCurl and XML.

3.1 Text Classification

Text training and classification began after the completion of data pre-processing. In this process, there was much software that could be used for text classification; among them, the commonly used sentiment tools included WEKA, WordFeature and ROST, etc. Regarding WordFeature that is simple, not only is its stability very weak, but the batch processing capacity of it should be further improved; for WEKA, it is less maneuverable. In this paper, ROST as a sentiment analysis tool developed by the virtual learning team of Wuhan University was selected to carry out text classification. During concrete operations, thematic information was extracted from Guba and then generated into a text set in a format required by ROST. Only one piece of information content to be analyzed was stored in each line. After analysis by ROST, serial number and the corresponding sentiment value were read for each piece of information to be analyzed from “Details of Sentiment Analysis Result”. Furthermore, such a result was just the totaling of positive, negative and neutral sentiment results, which conformed to the design thought of this paper.

3.2 Sentiment Scoring

After text segmentation has been realized by going through the above several procedures, what is more important is to perform sentiment scoring for words. Chinese words can be divided into three categories according to positive and negative attributes of sentiment, such as positive sentiment words, neutral sentiment words and negative sentiment words. In detail, positive or negative sentiment words can be distinguished by effectiveness of positive or negative sentiment respectively. Therefore, words in the lexicon were endowed with certain weights manually and affective intensity of each words ranges from 50 and -50. Considering that each piece of thematic information extracted from Guba may contain one or several of positive, neutral and negative sentiment words, scores of sentiment associated with a theme were directly added up at the time of computing sentiment scores of such a theme as the dimension adopted for positive, neutral and negative sentiment scoring was consistent. As shown in the corresponding formula, sentiment value of a theme=positive sentiment value of this theme + neutral sentiment value of this theme + negative sentiment value of this theme. After sentiment scoring has been conducted for data, values of all three categories of the sentiment in a certain period of time were added up to construct a comprehensive index. Network sentiment aggregation index at time t was denoted by \( X_t \), that is, \( X_t = \sum e x^e_t \), where \( e \) is the quantity of network sentiment information of \{buy, hold, sell\}. On this basis, Chinese investor network public opinion index can be expressed in,

\[
ISI_t = \log_2 \frac{1 + x^\text{buy}_t}{1 + x^\text{sell}_t}. \tag{1}
\]

4. Study on Investor Network Public Opinion Fluctuation Evolvement of Investors in China

4.1 Research Design

In general, intensive fluctuation of public opinions usually signifies high click ratio and search volume, and continuous reprinting and updating of information contents of the hot issue. However, status or tendency of public opinions cannot be clearly defined if only the search volume or click ratio
is taken into account. For this reason, investor network public opinion status and search volume were combined together in this paper to study fluctuation evolvement of network public opinions. Network data extracted from Guba from 2015 to 2016 were used as the test data.

In the first place, information of the number of web pages associated with a network platform Guba was collected by a search engine Baidu in such a time slot. Subsequently, such data information was totaled up to highlight abrupt change information of webpage quantity and lower the impact of limitations of a single search engine technology on exponential effects. 557 sets of data were acquired in total, which was denoted as a sequence of \( \{x_t\} \) and \( \{\hat{k}_t\} \). After transform, \( k_t = (x_t - x_{t-1})/x_{t-1} \) was obtained, where \( x_t \) refers to the search volume at time \( t \) and \( \hat{k}_t \) to search changing rate at time \( t \). In addition, time series after transform contained 556 pieces of information and the maximum value among them was 158.2%. For the convenience of later analysis, a constant was added to each piece of information uniformly (it was set as 100% here) to make all data positive. In this way, a new sequence of changing rate could be achieved and it was denoted by \( \{Y_t\} \) (\( i = 1, 2, \cdots, 556 \)).

To sum up, the comprehensive index for Chinese investor network public opinions is written into \( NP_{0_t} = ISI_t + Y_t \).

4.2 EGARCH Model Establishment

Through fundamental analysis and significance testing of the sequence, residual series of a random walk model was determined to have an ARCH effect. As nonnegativity restriction of quotient parameters from the GARCH model limited dynamism of conditional heteroskedasticity, etc., an EGARCH model was established to preferably describe features of data related.

In this paper, four models of EGARCH (1, 1), EGARCH (1, 2), EGARCH (2, 1), EGARCH (2, 2) were adopted to perform comparison. According to parameter significance testing after parameter estimations of such four models, models that failed such testing were eliminated. Among them, the model with the maximum likelihood value and the minimum AIC, SC and criterion value was selected as the final model. If these four models still cannot meet the corresponding conditions, order extrapolated based on items of EGARCH should be further expanded.

\[
\begin{array}{|c|c|c|c|}
\hline
 & \text{EGARCH (1, 1)} & \text{EGARCH (1, 2)} & \text{EGARCH (2, 1)} & \text{EGARCH (1, 2)} \\
\hline
\text{AIC} & -0.614008 & -0.776502 & -0.687456 & -0.570987 \\
\text{SC} & -0.601123 & -0.693233 & -0.517895 & -0.421987 \\
\text{Log} & -0.604776 & 74.05048 & 62.77584 & 52.11235 \\
\text{R}^2 & -0.652345 & 0.333464 & 0.192345 & 0.223497 \\
\hline
\end{array}
\]

By comparing results given by four models in the table above, EGARCH (1, 2) was selected as the optimal model. Parameter estimation results of this model were as follows.

\[
\text{Mean equation: } l_t = 0.7094 + 0.1612l_{t-1} + \varepsilon_t \tag{2}
\]

\[
\text{Variance equation: } \ln(h_t) = -0.0382 + 0.4235 \left| \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| - 0.5482 \left| \frac{\varepsilon_{t-2}}{\sqrt{h_{t-2}}} \right| - 0.1215 \left| \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| + 0.8706 \ln(h_{t-1}) \tag{3}
\]

In EGARCH (1, 2) model, a coefficient denoted as \( \gamma \) reflecting leverage effect was below 0 (\( \gamma < 0 \)), which indicated that the data had an asymmetric influence on external shock. In the case of \( \varepsilon_{t-1} > 0 \), this information shock applied a shock of 0.302 times on logarithm of conditional variance; by contrast, if \( \varepsilon_{t-1} < 0 \), its shock on such a logarithm turned into 0.545 times. Analysis described here conformed to the conclusion drawn above. Such a result signifies that fluctuations caused by negative news have a leverage effect during the diffusion of public opinion information. In other words, it is much likely for Chinese investor network public opinion fluctuations asymmetrical to a certain extent to suffer adverse impacts.

Volatility of investor network public opinion information evolvement in China was investigated in this paper. Firstly, an index was constructed for investor network public opinions in China by
network information acquisition and analysis. On this basis, a comprehensive index was further established for public opinion fluctuations in combination with search volume and click ratio. Additionally, it was proposed to build an EGARCH model depending on relevant volatility studies in financial field to perform public opinion information volatility fitting. As demonstrated by fitting results, fluctuations of investor network public opinions exhibited some characteristics of volatility clustering and asymmetry, etc. Consequently, they were more susceptible to impacts of negative news. Considering this, supervision departments shall take features of investor public opinion fluctuations in China into full account at the time of formulating related policies; as for important negative information spread on the Internet, it shall be clarified by the department concerned in a timely manner to guide investor’s expectations reasonably, calm down tendencies of public opinions and avoid the occurrence of unexpected group events that may severely affect stability of the financial system.

5. Conclusions
Volatility of investor network public opinion information evolvement in China was investigated in this paper. Firstly, an index was constructed for investor network public opinions in China by network information acquisition and analysis. On this basis, a comprehensive index was further established for public opinion fluctuations in combination with search volume and click ratio. Additionally, it was proposed to build an EGARCH model depending on relevant volatility studies in financial field to perform public opinion information volatility fitting. As demonstrated by fitting results, fluctuations of investor network public opinions exhibited some characteristics of volatility clustering and asymmetry, etc. Consequently, they were more susceptible to impacts of negative news. Considering this, supervision departments shall take features of investor public opinion fluctuations in China into full account at the time of formulating related policies; as for important negative information spread on the Internet, it shall be clarified by the department concerned in a timely manner to guide investor’s expectations reasonably, calm down tendencies of public opinions and avoid the occurrence of unexpected group events that may severely affect stability of the financial system.

References