Ca' Foscari
University
of Venice

Corso di Laurea Magistrale (LM) in
ECONOMIA E FINANZA

Final Thesis

How investor sentiment affects the risk of stocks?

Supervisor
Carlo Romano Marcello Alessandro Santagiustina

Graduand
Tianshu Zhang
Matriculation Number 883011

Academic Year
2021 / 2022

Ca' Foscari
University
of Venice

How investor sentiment affects the risk of stocks?

Ca' Foscari University of Venice

Abstract

Ca' Foscari University of Venice

With the advent of the information age, the Internet has played an increasingly important role in our lives. We use the Internet for information dissemination, gaming, and entertainment. It has led to a massive amount of information on the Internet every day. Many scholars have found that there is a specific correlation between network information and the securities market. Based on the scholars' research, this paper further analyzes how investor sentiment in online information affects the risk of individual stocks and how emotions affect the risk of individual stocks?

The object of this study is the investor sentiment represented by the information on the stock forum and the 50 constituent stocks corresponding to the Shanghai Composite 50 Index. The specific approach is to use self-written code to collect data on all postings of the 50 constituent stocks of the Shanghai Composite Index over the years 2015-2017, and use appropriate algorithms to divide posts into four categories named "positive," "negative," "Flat" and "excluded." Its corresponding browse volume assigns each post its weight, and a variable of investor sentiment index is constructed. Then, using this indicator, the quantile regression is performed on the log yields of the 50 stocks and the Shanghai 50 Index; respectively, the relationship between investor sentiment and VaR is observed.

Empirical results: VaR will increase when investor sentiment rises, and VaR will decrease when investor sentiment deteriorates. At the same time, investor sentiment has lagged effects on VaR, and investors' short-term and long-term effects on VaR are different.

This article's innovation: Before this, the academic community mainly discussed the relationship between investor sentiment and the stock market but did not discuss the risk relationship. This article will discuss the risk relationship and further explore the reasons for promoting or inhibiting risk.

Keywords: network information; investor sentiment; quantile regression; VaR

Ca' Foscari
University
of Venice

1. Introduction

1.1 Research background and significance

So far, the Internet has become important for the public to obtain information, and massive amounts of information and data are generated on the Internet every day. Many scholars have set their sights on Internet information, how to effectively use this information in market forecasts, explore the relationship between various types of information on the Internet and the market, and provide a basis for market analysis and investment decision-making' research.

With the emergence of more and more social media in China, investors have gradually obtained information through the Internet and shared information. Investors use the Internet to send or obtain a large amount of soft information related to listed companies, such as performance evaluations, bonus announcements, etc. With the increasing popularity of platforms such as Tieba and Weibo, the amount of information is increasing, and the speed of information dissemination is accelerating. We can guess that with the vigorous development of information technology, the entire stock market will be more and more affected by Internet information. Therefore, our study of the impact of emotional factors from Internet information on the stock market will become a meaningful subject.

Chinese scholars focused more on theoretical research, and there are relatively few empirical studies about investor sentiment. In the process of empirical evidence, constructing the index to measure investor sentiment correctly is a crucial step. On this basis, this paper divides the internet investor sentiment research into two categories: the first category, mainly to investors attention, posting volume, and other directly available digital information as a variable of Internet information, to study the impact of Internet information on stock prices. This kind of research only stays at direct access to digital information and fails to dig deeper into the information contained in the network text. Second, with the development of text classification technology in recent years, some scholars have begun to use text mining technology to obtain emotional information from the network information text, build investor sentiment indicators based on the information obtained, and analyze the relationship between emotional indicators and securities market variables with other directly available variables of the network.

1.1.1 Research on the direct quantification of network information

Wysocki's (1999) article first explored the relationship between Internet information and the stock market. Using the maximum number of posts from 50 companies from January to August 1998 as empirical data, this paper constructs the investor sentiment index and then finds some correlation between the next day's stock trading volume, abnormal yield, and investor sentiment index the day.

In addition to the number of posts, some scholars take the forum's scoring function as a variable to measure network information. Tumarkin and Whitelaw (2001) studied 72 stocks in the Internet industry, using the rating function of the famous U.S. forum RangingBull.com (the sender can make a buy, hold, sell evaluation of the stock) as a variable to measure the text information of the network, and found that in the higher comment information trading day, the volatility of investors' views was related to the abnormal industrial adjustment income.

PAULC (2007) completed the collection of Abbyst of The Market column review data for the Wall Street Journal between 1984 and 1999, quantitatively analyzing the correlation between stock earnings and online public opinion. PAULC (2007) concluded that if a large number of media predicted a downward trend in stock market returns, there would be a degree of rebound in equity fundamentals. The high volume of the stock market corresponds to the emotion of being too high or too low. The conclusion is that noise traders and liquid traders will bring about a downward trend in stock market returns under the influence of negative public opinion.

Chinese scholars Zhao Wei and Liang Yi (2009) used the number of Internet financial information bars to represent the intensity of information flow and studied the relationship between stock market yield fluctuations and Internet financial information volume. The article concludes that as the number of information increases, the return on stock increases, and when the amount of relevant information searched on the Internet increases several times, the fluctuation of stock yield will be disturbed by random factors and be significantly influenced by the amount of information. Another Chinese scholar, Rao Youlei and Wang Pan (2010), did a similar study: they analyzed the level of media attention to stocks (quantifying the number of news about listed companies) and found that the higher the media's interest in the more than 200 newly issued stocks, the higher the price of the shares. But in the long run, the impact of media attention on new shares is negative.

1.1.2 Text mining tool to quantify the network text information

The advantage of using forum posting volume and news comments to construct investor sentiment index is that the number of posts and comments can be presented digitally, which does not need to be quantified in particular, but the disadvantage is that the text in the post commenting on stocks is not analyzed. Some scholars abroad realized the importance of text analysis, so they began to use text mining tools to process text information as a way to quantify investor sentiment.

Feldman and Dagan (1995) summarized a KDT (knowledge Discovery In-Text) model based on KDD, one of the earliest text mining models. They were using the model, the mining of more than 22,000 financial-related stories published on Reuters, and successfully drawing a picture of the distribution of stock market performance with keywords such as country and organization.

Schumaker and Chen (2006) first quantify the stock market news text from words, sentences and chapters, and use the support vector machine model to predict the price trend of the stock market 20 minutes after the release of the Internet news; the results show that the prediction effect of the model with news variables is better than that of the traditional linear model.

Das and Chen (2007) considered another measurement method, using the forum information to get small investors' optimism about stocks to make relevant stock forecasts. They used the Bayesian classifier to calculate the frequency of words for information in Yahoo's financial forums, using the basis for quantifying investor sentiment. They concluded that there was a correlation between the commenting behavior of individual investors and the volatility of the stock market, which was not apparent at the micro-stock level, but at the macro-market level, the positive correlation between this market index and the investor's emotional inclination index was quite evident.

Tetlock (2008) taps into the mood of the news media by reporting on a Wall Street Journal column, using a simple method of calculating the frequency of words of optimism or pessimism, dividing news texts into positive and negative categories, and then analyzing the relationship between media sentiment and stock price movements and market volume. Media pessimism lowers the company's earnings expectations, but share prices don't fluctuate sharply with negative news, the study found.

Although the traditional text mining method is feasible, the disadvantage is that the amount of mining is small, and the text analysis is not thorough. Because of this, many scholars began to use more accurate third-party text mining tools to quantify indicators. Mao and Bollen (2011) collected nearly 10 million emotional statements on Twitter from February to December 2008 about the Dow, using Google's OptinionFinder and GPS tools. The use of these tools significantly improves the accuracy and efficiency of text mining. Finally, the author tested the correlation between the Dow Jones Industrial Average and the emotional value and found that the stable emotional value was significantly related to the Dow Jones Industrial Average, and when the stock market was relatively stable, the correct rate of predicting the next day's rise and fall of the Dow Jones Index was more than 87%.

It can be seen that more and more scholars study the subject. At first, scholars discovered a correlation between Internet information and stock market yields; The study found that the performance of stock market turnover was best predicted using directly observable Internet values, and the prediction model with text mining technology was better than the traditional model.

Scholars mainly use directly observable variables and variables constructed after obtaining information through text mining technology to characterize investor sentiment, the main content of the initial research is the correlation between investor sentiment and the stock market analysis, that is, whether it is a relevant, positive, or negative correlation. At present, the general conclusion is that it is relevant, and the influence may be different at different times and on different occasions. Further, the study adds the use of investor sentiment variables to forecast the stock market, including the prediction of trading volume, yield forecast, etc., in which the forecast effect of volume is better, the prediction of yield varies according to the different variable selection and the different accuracy of the occasion. It can be seen that many scholars have studied this problem and obtained many results, but the content of the study is relatively single; there is room for further excavation.
Using financial information on the Internet to study stock market volatility is a relatively novel method; it provides us with a new angle to analyze the stock market. And this method covers the theoretical results and measurement methods in many fields of stock research: machine learning (text mining) method, stock market forecast model, econometric theory, and technology. But in general, there are still some problems to be solved in this new stage of research, mainly the following two points:

1. With the progress of text mining technology, more and more scholars use text mining technology to process the network information text, more efficient access to information, but the processing method is still relatively rough because, in addition to the content of the comment itself, there is some other information of reference value. For example, the number of responses to ideas, the number of views viewed online, and so on.

2. At present, scholars have confirmed that there is a correlation between investor sentiment and the stock market in the network public opinion, and many scholars have studied the accuracy of investor sentiment on stock market forecast, but the research content is relatively single. Mainstream research is still stuck in the forecast of the stock market, although some scholars in

the study of the forecast yield, the use of investor sentiment variables to predict the volatility of stock prices, but no further exploration, no other portrayal of such volatility as risk, not even discuss how long such fluctuations last, let alone how investor sentiment caused volatility.

Because of the above problems, this paper is based on the research of scholars and further explores the impact of investor sentiment on stock risk.

1.1.3 The purpose and significance of the study

The purpose of this paper is to study the time and extent of the impact of investor sentiment on China's stock market risk reflected in Internet information. The research of this topic needs the knowledge of economics and computer science, especially the text analysis technology in the field of computers and the Python program, which provide the most basic data support and index quantification method for the empirical part of this paper. The primary purpose of this paper is to analyze the duration and degree of impact of Internet information on China's stock market risk, which can be summed up as the following questions:

(1)Does investor sentiment affect stock returns?
(2) How long is the impact of investor sentiment on individual stock risk?
(3) How does investor sentiment affect individual stock risk?

The above problems are the primary research purposes of this paper. At the same time, if the expected effect of this paper is good, the text mining can be a good quantification of investor sentiment, such research methods can be further expanded, such as the analysis of the financial statements Chinese the word part of the valuable soft information, to study the type of financial statements of listed companies on the impact of stock price fluctuations of listed companies. Therefore, the research ideas of this paper can continue to be used in other interdisciplinary issues, with a particular reference value.

1.2 Research ideas and methods

This article integrates text mining technology in the computer field ,mainly the financial dictionary method established by Ruan Jin, Yuan Jingrui, and Liang Xun(2008) and quantile regression in financial metrology to analyze how investor sentiment factors in online text affect individual stock risk. Namely, bullish sentiment, bearish sentiment, and flat sentiment respectively promote or inhibit VaR? How long is the impact? How does it affect VaR?

Ca' Foscari University of Venice

This article mainly focuses on the data of the DONGFANG stock forum(guba.com.cn), analyzes its sentiment through the method of text mining, quantifies it as an indicator of investor sentiment, and studies the relationship between investor sentiment and the SSE 50 Index and the VaR of 50 constituent stocks. The research idea is to first use text mining technology to mine the investor sentiment information contained in the stock forum' s posts, turn it into an investor sentiment indicator, quantify it into a numerical value, and then use the econometric quantile regression analysis method to analyze the investor sentiment The quantile that affects the logarithmic return of the Shanghai 50 Index and the 50 constituent stocks, that is, the effect of investor sentiment on the risk of individual stocks, and the duration of the impact. The research on these issues has filled some gaps in the current systematic analysis of the relationship between network information and stock market volatility. It is of great significance both in theory and in practical applications.

The logical framework for this article to analyze the relationship between Chinese stock market risk and investor sentiment on the Internet is as follows: (1) There is no direct digital information in the text in Tieba, so this article uses text mining method to compare the posts in Tieba. Perform word segmentation analysis and use a specific model to quantify it as an indicator of investor sentiment. (2) Combine the investor sentiment index after quantification of the text information of the stock bar post with the SSE 50 Index and 50 constituent stocks, and use the quantile regression method of econometrics to analyze how investor sentiment affects stock market risk and influence The length of time and other issues. The specific implementation of the technology is as follows:

1. Data collection. DONGFANG is the most visited and influential financial portal website in China and even the world. The information on its stock bar forum is the most relevant to the stock market, so the DONGFANG stock bar forum is selected as the source of online public opinion data. Use a self-written Python web crawler to collect all the posting data of the 50 constituent stocks of the Shanghai Stock Exchange 50 Index in the last two years, including posting time, title, page views, etc. Import the collected data into a CSV file, and then preprocess the data through Python. At the same time, download the closing price data of the Shanghai Stock Exchange 50 Index.
2. The classification of the post. First, establish a financial dictionary to classify the words that may appear in the posts, which are divided into four categories: "positive," "negative," "equal," and "excluded." Such as: "accumulate", "buy", "buy bottom", "speculate bottom", "good", "strong", "up", "flying red", "burst", "pull up", "increased position", Words such as "long", "covering", "bull market", "rebound", "pull up", "pull up", "hold" and other words are classified as "active", "reduce", "sell", " "Weak", "fall", "clearance", "cut meat", "short", "lighten up", "run", "quilt cover", "set dead", "escape", "floating green", "bear market", " Words such as "smashing", "losing", "diving", "withdrawing", "latching", "pressing" and other words are classified as

"negative", in order to ensure that the appearance of positive and negative emotion words in the post can correctly express investor sentiment , This article sets up "exclusion" words, such as: "no", "?", "no", "no", "how", "may" and other words that express negative and questioning tone. In other words, once these words appear in a post, then skip the post and not collect and analyze the investor sentiment of the post. This is because there are many negative and ironic posts in the stock bar posts, and simply extracting the positive and negative words from them cannot correctly reflect the sentiment value in the post. Because in actual operation, the posts in the stock bar alone contain a small number of words, and the number of posts themselves is large, so removing these posts that easily affect the results will not have a significant negative impact on the results. Excluding the above three types of posts, the remaining posts are all classified as "neutral."

From the extracted post content, the three types of words "positive," "neutral," and "negative" are assigned 1, 0, and -1, respectively. If a post contains two positive words, the independent sentiment value of the post is 2.

3. Calculation of indicators. According to the classification results, the calculation formula of investor sentiment indicators is constructed to calculate the specific numerical sequence of investor sentiment indicators for each day of the two years from the last two years, forming a string of time series.

4. Quantile regression analysis. Investor sentiment indicators are used to establish a quantile regression model with the SSE 50 Index and the log return of 50 constituent stocks (hereinafter referred to as the yield), in which the investor sentiment indicator is the dependent variable. According to the establishment of different models, analyze the impact of investor sentiment on stock market risks.

1.3 Innovation

This article focuses on the impact of investor sentiment on individual stock risks. This research has the following innovations:

(1) This article uses the investor sentiment indicators constructed by the previous scholars to conduct a more in-depth analysis based on Jin Xuejun(2013)'s conclusion that there is a correlation between DONGFANG stock returns. This article uses quantile regression to analyze the degree of influence of investor sentiment on individual stock risk and the influence cycle. It is the first time to analyze individual stock risk by combining online text information with stock quantitative risk indicators. This will fill the gaps in academic research in this area and make an essential contribution to the improvement of the theoretical system of the influence of network information on the securities market.

2. Empirical design

2.1 Text mining, Chinese word segmentation technology

How to quantify these investor sentiments and construct investor sentiment indicators in the network of investor sentiment information collected in this article and the text of Tieba requires the use of text mining technology. The smallest semantic unit in Chinese is a word. Chinese word segmentation refers to the process of segmenting a series of Chinese characters into meaningful multiple words according to a specific rule. The effect of Chinese word segmentation directly affects the effect of text mining.

There are currently three Chinese word segmentation algorithms. The first is based on understanding to segment the text, but the disadvantage is that it is time-consuming and labor-intensive and is not suitable for the analysis process of a massive amount of data. The second is to segment words based on statistics. The third is to segment words based on string matching. This article uses the third method of word segmentation. The basic idea of this method is to first establish a dictionary related to its analysis indicators and then compare the entries in the text with the entries in the dictionary. Once they are consistent with each other, then Separate the term from the original string, and then re-comparison until the final word segmentation is completed.

2.2 Quantile regression

This article uses VaR to measure the risk of the stock market and uses the Quantile Regression model. The following introduces the model proposed by Koenker and Bassett (1978):
For a univariate distribution function $F_l(x)$ and a given probability q, and 0<q<1, it is said:

$$x_q = \inf\{x|F_l(x) \geq q\}$$

Q quantile for $F_l(x)$.
For a given probability q=1-p, $\{x_t\}$, the q quantile can be obtained by :

$$\hat{x}_q = argmin_{\beta} \sum_{i=1}^{n} w_q(x_i - \beta' z_t)$$

Which $w_q(z)$ is defined as :

$$w_q(z) = \begin{cases} \text{qz} & \text{z} \geq 0 \\ (q-1)z & z < 0 \end{cases}$$

Then we define the quantile regression model as:

$$x_t = \beta' z_t + \varepsilon_t$$

The conditional quantile function is:

$$\text{Q}(\text{q}|z_t) = \hat{\beta}' z_t$$

2.3Investor sentiment indicator construction

In the process of studying the relationship between investor sentiment and stock market risk, this thesis not only uses text mining technology and quantile regression in econometrics but also constructs an indicator, the investor sentiment indicator.

This thesis uses the investor sentiment indicator constructed by Lin Quansheng (2016). The mathematical definition of investor sentiment indicator $B_t$as as follows:

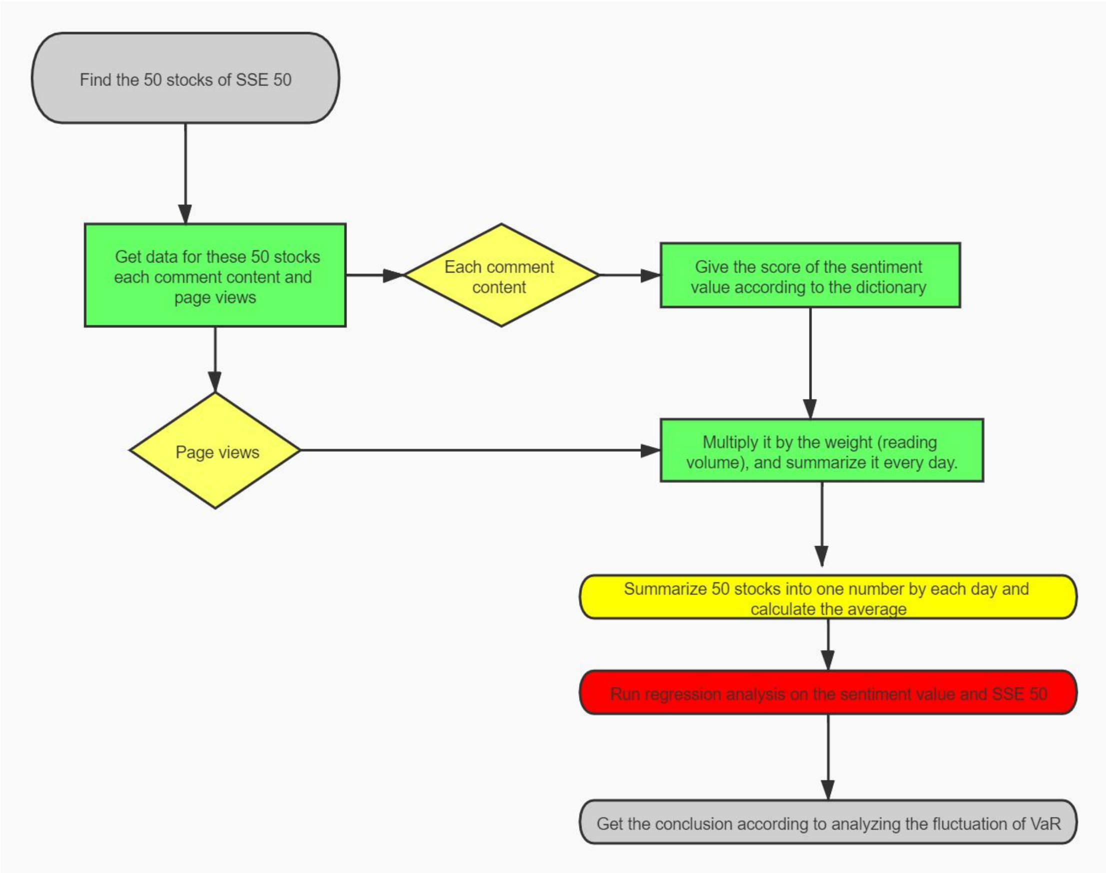
$$B_t = ln\frac{(1 + M_t^{Buy})}{(1 + M_t^{Sell})}$$

$$M_t^C = \sum_{i \in D_t} w_i x_i^C$$

$$w_i = \frac{a_i}{\sum_{i \in D_t} a_i}$$

Among them, C{negative (sell), neutral (hold), positive (buy)}; x_i^C the value is 1 or 0, if the post I belongs to category C, then x_i^C =1, otherwise x_i^C =0; the weight is an innovative point of this article. The attention of each post is different, and it is not a good idea to attach the same attention weight to all posts. We noticed that the amount of visits to each post is different, and the number of visits can reflect the degree of investor attention to the post and the degree of influence the post has caused. Therefore, this article uses post attention a_i as the weight of each post.

3. Empirical analysis



3.1 Data sources

In this paper, the Shanghai Stock 50 Index, its constituent stocks, is taken as an example to analyze the relationship between online public opinion and securities market risks. DONGFANG is the most visited and influential financial portal website in China and even in the world, so this article uses DONGFANG Stock Forum as the source of data. This article uses a self-written Python program to collect all the posting data of the 50 constituent stocks of the Shanghai Stock Exchange 50 Index of DONGFANG Finance and Economics in the years 2015-2017 and uses related models to quantify it as an indicator of investor sentiment.

The data of the stock market variables used in the relevant analysis in this article are the closing prices of the Shanghai 50 Index and 50 constituent stocks from the years 2015-2017, and the logarithmic return rate is further calculated. Therefore, this article finally uses the investor sentiment indicator, the SSE 50 Index, and the logarithmic rate of return of 50 constituent stocks for research.

3.2 Empirical analysis

This part is about the quantile regression method to study the influence of investor sentiment and individual stock risk.

After obtaining the investor sentiment index, we apply the quantile regression method to study the relationship between investor sentiment and individual stock risk. Consider the simplest model $Q(q|B_t) = \alpha + \beta B_n$. Under the VaR of 0.90 and 0.95, quantile regression is performed on the Shanghai 50 Index and 50 constituent stocks, respectively.

If the regression result of the constant term and the investor sentiment index is very significant, it indicates that there is a clear correlation between investor sentiment and individual stock risk. The VaR value generally takes the negative value of the quantile to indicate the size of the loss. Therefore, if the $\beta < 0$, it means that the higher the investor sentiment in the market, the greater the risk of individual stocks.

tau:0.90	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	0.01039	0.00124	8.40757	0.00000
score	−0.29452	0.12617	−2.33437	0.01998

tau:0.95	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	0.01546	0.00227	6.82286	0.00000
score	−0.33179	0.17416	−1.90512	0.05736

The above regression results show that the regression results of the constant term and the investor sentiment index are very significant. When the quantile is set to 0.90, the result is significant at the 95% confidence level, and when the quantile is set to 0.95, the result is at 90%. The confidence level is significant. From the above regression results, the conclusion is basically in line with expectations; that is, the higher the investor sentiment in the market, the greater the risk of individual stocks.

Next, consider the impact of investor sentiment on the risks of individual stocks with different maturities. At a quantile of 0.9, the P-value of the estimated investor sentiment coefficient in the simplest model is the smallest. Consider the investor sentiment seven days (one week) and 20 days (one month) in advance for quantile regression. Observe the size of the coefficient to determine whether investor sentiment has a more significant impact on long-term risk.

tau=0.9, tag=7	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	0.00933	0.00131	7.13444	0.00000
score	0.00718	0.04714	0.15221	0.87909

tau=0.9, tag=20	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	0.01017	0.00128	7.96051	0.00000
score	−0.02937	0.06434	−0.45637	0.64833

The above regression results show that the influence of investor sentiment on long-term risk is not significant. And with the shorter the lag period, the more obvious the influence of investor sentiment on risk.

4.conclusion

Whether Internet financial information can affect stock prices in China's securities market is one of the most popular frontier financial issues in recent years. Scholars' research is mainly focused on constructing investor sentiment indicators to quantify investor sentiment in online information to predict stock prices. This article uses the investor sentiment indicators constructed by scholars to measure online investor sentiment and further analyzes the impact of investor sentiment on individual stocks.

This article combines online investor sentiment with individual stock risk analysis. This study uses the quantile regression of the Shanghai 50 Index and the 50 constituent stocks' logarithmic returns and investor sentiment to analyze: (1) What is the difference between the impact of investor sentiment on the VaR of individual stocks under different confidence levels? (2) What are the impacts of investor sentiment of different maturities on individual stock risks? The short-term or long-term impact is significant. (3) Consider the short-term and long-term impacts at the same time and whether the impact directions are the same.

The innovative point of this article is that it focuses on the relationship between online investor sentiment and individual stock risk for the first time and analyzes the reasons for the impact for the first time. This will enrich the research content to a large extent, fill the gaps in current academic research, and provide an empirical basis for follow-up research.

It is true that this article still has a considerable degree of limitations and deficiencies that need to be improved in the future:

First, the DONGFANG data of the stock bar posts used to construct investor sentiment indicators is noisy. Although the content of DONGFANG is close to the stock market, there are still posts that are less relevant to the stock market. At the same time, the quality of some posts cannot be guaranteed, and there are suspicions of useless stickers. All of these have brought significant interference to the construction of sentiment indicators. How to accurately remove interference samples while avoiding the loss of precious effective samples is an improvement direction in the future.

According to the Jia Li, Yun Chen, Yan Shen(2018). Measure China's stock market sentiment when the text sentiment is abnormally high or low, the transaction volume will be higher, and when there are more disagreements, it is measured by our text disagreement. Based on a large number of text data sets, our analysis is noise trading theory and arbitrage limitation theory, as well as prediction models from limited attention and divergence. This can avoid the influence of noise to a certain extent.

Second, text mining technology may not be able to classify accurately. At present, the Chinese word segmentation thesaurus still needs to be improved, and the use of the Chinese word segmentation thesaurus for word segmentation may not be wholly accurate and successful.

Third, the sample data studied in this article is still tiny. This article only studies the 50 constituent stocks of the Shanghai 50 Index. Compared with the number of stocks owned by the entire market, it is relatively small. The 50 constituent stocks cannot fully represent all individual stocks. Therefore, the sample size can be increased in the future, and further exploration can be carried out.

Fourth, we can add some other variables, hot vocabulary, etc., for more extensive consideration, such as the Baidu index. Baidu is the largest Chinese search engine, similar to Google. The Baidu Index is a free massive data analysis service based on Baidu web search and Baidu news search to reflect the "user attention" and "media attention." of different keywords in the past period of time. According to the article Jia ChunFang, Giray Gozgor, Chi-Keung Marco Lau, Zhou lu(2020). The impact of Baidu Index sentiment on the volatility of China's stock markets, adding Baidu Index to the construction of sentiment indicators can more effectively improve the prediction performance, especially in Periods of high volatility, because in periods of market turbulence, investors tend to seek additional market information to avoid risks and increase search volume.

The quantile autoregressive model can be used to perform regression of this sentiment variable index in units of seasons, months, and weeks, which can expand the index to a wider range of applications; conversely, it can also make the index more practical.

Research directions that can be further expanded: use more effective word segmentation methods and more samples to conduct another analysis; whether the conclusions are consistent in different regions and different industries in different seasons, etc.

Reference

[1]Antweiler,Murray Frank,2004.Is all that talk just noise? The information content of internet stock message boards.[J], Journal of Finance.Vol59,No.3,pp.1259-1294.

[2]Schumaker,R.P. and Chen,H.,2006, "Textual Analysis of Stock Market Prediction Using Financial News," Americas Conference on Information Systems.

[3]Das, S.& Chen, M. Yahoo! for Amazon: Sentiment Parsing from Small Talk on the Web, Management Science, vol.53, pp.1375-1388,2007.

[4]Tetlock,P.,Saar-Tseachansky,M.Mackay, S.More Than Words: Quantifying Language to Measure Firms' Fundamentals.Journal of Finance,
VOL.LXIII,no.3,pp.1437-1467,2008.

[5]Cortes, Corinna; Vladimir Vapnik(1995)." Support-Vector Networks." Machine Learning.20:273-297.

[6]Koenker RW,Bassett GW.Regression quantiles.Econometrica 1978;46:33-50.

[7]Granger, C.W.J. 1969， Investigating causal relations by econometric models and cross-spectral methods.Econometrica,37,424-438.

[8]Wysocki, P.D.Cheap talk on the Web: The determinants of postings on stock message boards[J].Ann Arbor,1998,1001:48109-1234.

[9]Tumarkin,R.,Whitelaw,R.News or Noise? Internet Postings and Stock

Prices[J].Financial Analysts Journal,2001,57(3):41-51.

[10]PAULC, Giving content to investor sentiment. The role of media in the stock market:

The Journal of Finance;2007(06).

[11]Jia Li,Yun Chen,Yan Shen.Measure China's stock market sentiment;2018

[12]Jia ChunFang, Giray Gozgor,Chi-Keung Marco lau,Zhou lu.The impact of Baidu Index sentiment on the volatility of China's stock markets;2020(01)

[13] Ruan Jin, Yuan Jingrui, Liang Xun, 2008, "Research on Text Weight Reduction Method of Internet Financial News Search," "Journal of Xihua University," "Finance and Finance,"

[14] Hu Lingyun, Hu Guilan, Xu Yong, Li Longshu, 2010, "Research on Web-based News Text Classification Technology," "Journal of Anhui University," 34(6)

[15]Jin Xuejun, Zhu Yu, Yang Xiaolan. The influence of online media on the stock market: an empirical study of Oriental Fortune Internet stock bar [J]. Journalism and Communication Research, 2013.

[16] Lin Quansheng, Research on the Relevance of Internet Information and Stock Market—Analysis of Text Mining Based on Stock Forum. Chengdu: the Southwestern University of Finance and Economics 2016.

[17] Cold Winter, Procyclicality Study of VaR Model Measuring Market Risk. Hefei: University of Science and Technology of China 2014.

[18] Liu Huangsong, The Theory of Intrinsic Investment Value of Stocks and the Problem of China's Stock Market Bubble[J], Economic Research, 2005.

[19] Xu Ainong, Research on the Measurement and Rationality of China's Stock Market Bubble [J], Finance and Economics Theory and Practice, 2007.

[20]Zhao Wei, Liang Dun. Research on the Volatility and Correlation of Internet Financial Information Yield. Computer Technology and Development, Volume 19, 2009.

[21] Rao Yulei, Wang Pan. The impact of media attention on the performance of new stocks: evidence from the Chinese stock market [J]: Accounting and Finance, 2010.

Ca' Foscari
University
of Venice

Thanks

First of all, I would like to thank my thesis supervisor, Professor Carlo Santagiustina. He made instructive opinions and recommendations on the research direction of my thesis. During the process of writing the thesis, he gave timely and careful guidance on the difficulties and doubts I encountered and put forward many helpful suggestions for improvement. In addition, Professor Carlo Santagiustina also asked entirely meaningful questions about the logical framework in the first draft of my paper and helped me sort out the logic of my paper, which significantly improved the logical system of this paper.

In addition, I would also like to thank my friends and classmates for their tremendous support and help in writing the paper, which gave me great inspiration. I would also like to thank the authors in the references. Through their research articles, I have a good starting point for the research topic.

Finally, I sincerely thank family, friends, and classmates. It is with their encouragement and support that I can complete this paper.