

The Impact of Social Media Sentiment on Stock Market Based on User Classification

Lin Liao^{a1} and Tao Huang^b

^aBusiness School, Beijing Wuzi University, China

^bGuanghua school of management, Peking university, China

Abstract. The relationship between social media sentiment and the stock market has been receiving much attention. Based on the perspective of social media user classification, using Weibo and Shanghai Composite Index data, quantile regression and instrumental variable quantile regression (IVQR) model is built to explore the impact of sentiment and sentiment fluctuation on authenticated and non-authenticated users on stock market returns respectively. The research results show that Weibo sentiment and sentiment fluctuation have a positive impact on stock market returns, but the effects of the two types of users are different. The sentiment of authenticated users has a stronger and longer impact on stock market returns, while only moderate sentiment and sentiment fluctuation of non-authenticated users have a positive impact on stock market returns. The research provides evidence for the relationship between social media sentiment and stock market, and has some practical significance for both social media platforms and users.

Keywords. Social media sentiment; Stock market returns; User classification; IVQR

1. Introduction

The prediction of the stock market is a matter of great interest to both the academic and business communities. Some behavioral finance research shows that the stock market is not completely consistent with the random walk theory [1]. There are still some predictable components that can affect the stock market, and investor sentiment is one of the important factors [2,3]. As a result, researchers are increasingly focusing on using investor sentiment to predict stock market behavior. Investor sentiment is defined as investor enthusiasm or negative for the future stock market [4]. Previous researchers have tried to adopt a series of proxy variables to reflect individual investor sentiment, in order to explore the relationship between investor sentiment and the stock market. These variables include consumer confidence levels [2], investor surveys, equity capital flows, closed-end fund discount rates, and asset buy / sell ratios [5,6].

With the high development of Web 2.0 and social networks, recent researchers have attempted to measure public sentiment with the help of big data on social media platforms [6,7]. Although these studies have made beneficial attempts on the relationship between social media sentiment and the stock market, there are two main limitations. The first is the lack of attention to the differences and classification of social media users. In social media, not all users have the same influence, and the sentiment of high-impact

¹ Corresponding Author, Lin Liao, Beijing Wuzi University, China; Email: liaolin@bwu.edu.cn

users and ordinary users may have different effects on the stock market. Previous studies have shown that [8], in terms of the dissemination of public opinion, opinion leaders' emotional publicity and ordinary Internet users' reasonable publicity are more communicative. Therefore, focusing on differences or classifications of social media users can explore the relationship between sentiment and the stock market more accurately. Second, most previous studies were conducted under the framework of mean regression, which can reveal the impact of investor sentiment on the average level of stock market performance, but cannot characterize the heterogeneous effects in different quantile ranges. In order to make full use of the characteristics of each quantile interval, the quantile regression method is increasingly applied to the financial field [9]. However, using traditional quantile regression methods can only reflect the correlation between social media sentiment and the stock market. Because the quantile level of stock market returns may affect the sentiment (generally expressed as the higher the quantile level of stock market returns, the more positive the sentiment), the endogenous nature of the model is caused. In order to solve the endogenous problem, this article introduces the sentiment in social media of the previous period as an instrumental variable and constructs the IVQR model [10] to quantify the effects of different social media users (authenticated and non-authenticated users) on the stock market.

2. Related Works

In recent years, with the popularity of social media applications on the Internet, researchers have begun to detect public sentiment through some mainstream social media sites and associate big data on the Internet with the stock market. Previous related studies used effective text analysis techniques to analyze social media. The measurement of emotions in social media can be classified into three categories. The first category is to directly use text analysis software or construct a classifier to calculate the positive and negative sentiment. For example, research by Zhang et al. [6] show that user pessimism contained in social media will cause stock prices to fall in the short term, and user pessimism is negatively correlated with stock market indexes. Huang et al. [7] generate time series of emotional tendency by capturing Sina Weibo data, and use Granger causality test and support vector machine model to prove that the changing tendency of high positive emotional tendency in Weibo can affect the change of the closing price of Shanghai Stock Exchange Index. In addition, adding Weibo sentiment information to the stock prediction model can increase prediction accuracy. Renault [11] bases on user-published content on StockTwits and builds a vocabulary used by online investors to measure sentiment polarity indicators. It is found that online investor sentiment can help predict stock returns for that day.

The second type of research indirectly reflects the negative or positive sentiment of investors through the number of specific words on social media. For example, Zhou et al. [12] calculate sentiment by selecting the number of daily occurrences of six specific hot words on Sina Weibo, including "Bull Market", "Bear Market", "Positive news", "Negative news", "Stock Index rise" and "Stock Index drop". Studies have shown that there is a Granger causal relationship between the number of keywords "Stock Index rise" and "Stock Index drop" and the closing price of the Shanghai Composite Index, and that "Stock Index drop" is more accurate than "Stock Index rise" in predicting the stock market. Cheng and Lin [13] divide the words expressing user sentiment into two types of polar words, "bullish" and "bearish". They construct the ups and downs index and find

the positive correlation between social media investor's ups and downs index and the index returns and trading volume of stock market.

The third category focuses on the divergence of sentiment. For example, Siganos et al. [14] argue that the greater the divergence of sentiment, the greater the difference in perspectives on expectations, risk, and stock value. By establishing a panel data model based on Facebook data and stock data from 20 countries, a positive correlation between sentiment divergence and uncertainty in trading volume and stock price is found. This relationship is stronger when individual investors are more likely to trade. Some researchers [15] use text mining technology to extract sentiment indicators from data obtained from Twitter, applying two methods of linear regression and support vector regression, and find that the individual's divergence of sentiment includes uncertainty about stock, and can be used to improve the accuracy of stock price prediction.

3. Theoretical Background

Traditional efficient market theory (EMH) divides securities markets into weakly efficient markets, semi-strongly efficient markets, and strongly efficient markets based on the amount of information investors can obtain on the securities market [16]. Behavioral finance theory breaks this division of the traditional market and incorporates investors' psychological factors into market analysis. The study of the impact of investor sentiment on investment decisions is mainly based on psychological research [17]. Research on behavioral finance theory finds that investors are not completely rational in the decision-making process under uncertain conditions, and will be affected by beliefs such as overconfidence, representativeness, availability and other systematic cognitive biases [18]. These judgment biases can be divided into three categories: empirical simplification rules, self-deception, and emotional basis judgments [19].

Among them, emotion basis judgment refers to the fact that in the decision-making process, people influence decision-making due to changes in their emotions. Personal sentiment is affected by group sentiment, so that the stock market cannot be analyzed objectively. Investor sentiment refers to a belief formed by investors based on their expectations of future cash flows and investment risks, but this belief does not fully reflect the current facts [4]. With limited information, investors cannot fully understand the information in the market. Therefore, when making investment decisions, they may not be completely rational, but they are prone to blindly follow the trend. This shows that investor sentiment will affect investment decisions and then the stock market.

The advent of the Web 2.0 era has provided more possibilities for the collection of investor sentiment data. The essence of Internet sentiment is the sentiment atmosphere. The sentiment of all users together constitutes the sentiment atmosphere in the network, and users will be infected and affected by the sentiment of others in the same network. With the development of the Internet, online social networking and its platform have been widely popularized and applied. Although the emotional information of Internet users is distributed in various corners of the Internet, it is mainly concentrated in online social platforms such as blogs, Twitter, and Facebook [20]. Among them, Sina Weibo (hereinafter referred to as Weibo) is currently the largest microblogging service social network in China. According to statistics released by Sina, as of June 2018, the number of Weibo users across the country reached 337 million, with an average of 120 million posts published daily. Based on the huge amount of data, Weibo information can reflect changes in the overall psychology and behavior of society accurately and promptly.

Therefore, Weibo is selected as the representative of social media platforms to conduct research, focusing on exploring the relationship between the sentiment of different types of Weibo users and the stock market.

4 Methodology

4.1 Quantile regression model

The quantile regression model was first proposed by Koenker and Bassett [21] to study the relationship between the different quantiles of the explanatory variables and the explanatory variables. Compared with ordinary mean regression, quantile regression can fully describe the effect of explanatory variables on the explanatory variables, and can capture the characteristics of the upper and lower ends of the distribution of the regression variables. Moreover, it is not necessary to assume the distribution of random perturbations in the model, which makes the model have a larger adaptable range and stronger robustness. The linear quantile regression model is:

$$Q_r(\tau|s) = \alpha(\tau) + s \cdot \beta(\tau) \quad (1)$$

Where $\alpha(\tau)$ is the intercept term, $\beta(\tau)$ is the slope coefficient to be taken into account, r is the stock market return, and s is the Weibo sentiment, $Q_r(\tau|s): \inf\{\xi: F_r(\xi|x) \geq \tau\}$ represents the τ -th quantile of the random variable r . The method of parameter estimation is:

$$(\hat{\alpha}(\tau)\hat{\beta}(\tau)) = \arg \min_{\alpha, \beta} \sum_{i=1}^n \rho_\tau(r_i - \alpha - s_i \beta) \quad (2)$$

Where $\rho_\tau(u) := u(\tau - I(u < 0))$ is the loss function.

According to the Skorohod theory, for the random variable r , there is a random variable $U \sim \text{Uniform}(0,1)$, so that r can be expressed as $r = Q(U)$. Where $Q(\tau)$ represents the τ -th quantile of r (Durrett, 2005). Therefore, the linear quantile regression model can also be expressed as:

$$r = \alpha(U) + s\beta(U), U|s \sim \text{Uniform}(0,1) \quad (3)$$

Where $\alpha(\tau) + s\beta(\tau)$ is a continuous and strictly increasing function of τ . When the value of the random variable U is τ , $\alpha(\tau) + s\beta(\tau)$ is the τ -th quantile of r under the known condition of s , and the corresponding regression coefficient $\beta(\tau)$ can be estimated using the traditional model estimation method.

4.2 Instrumental variable quantile regression model

Instrumental variable quantile regression (IVQR) was first proposed by Chernozhukov (10) to solve the endogenous problem in quantile regression models. In the model $r = \alpha(U) + s\beta(U)$, the sentiment s may be affected by the quantile level U of the stock market return r , that is, the higher the stock market return, the higher the investor's sentiment. The intuitive correlation between s and U will cause the endogenous problem of the model, resulting in the inconsistent parameter estimation by traditional quantile regression methods. Considering the influence of stock market returns on sentiment, this paper introduces instrumental variables to construct the IVQR model. The instrumental variable needs to be related to the independent variable s , but not affected by the

dependent variable r . Therefore, this article selects the sentiment z of the previous period as an instrumental variable, which has a strong correlation with the sentiment in current period, but will not be affected by the current stock market returns. After introducing z , the IVQR model is:

$$r = \alpha(U) + s\beta(U)$$

$$U|z \sim \text{Uniform}(0,1) \quad (4)$$

Where s is related to U , and z is an instrumental variable. The constructed quantile function (SQF) is:

$$S_r(\tau|s) = \alpha(\tau) + s\beta(\tau) \quad (5)$$

Since event $\{r \leq S_r(\tau|s)\}$ is equivalent to event $\{U \leq \tau\}$, we can get

$$P[r \leq S_r(\tau|s)|z] = P[U \leq \tau|z] = \tau \quad (6)$$

That is, $P[r - S_r(\tau|s)|z] = \tau$, which means the τ -th conditional quantile of the random variable $r - S_r(\tau|s)$ is 0. So there are:

$$\arg \min_{f \in F} E[\rho_\tau(r - S_r(\tau|s) - f(z))] = 0 \quad (7)$$

Then the problem is transformed into finding the appropriate $S_r(\tau|s)$, so that the above equation holds. In order to simplify the calculation, let $f(z) = z \cdot \gamma$, and $S_r(\tau|s) = \alpha(\tau) + s\beta(\tau)$, then the objective function can be expressed as:

$$Q_n(\tau, \alpha, \beta, \gamma) := \frac{1}{n} \sum_{i=1}^n \rho_\tau(r_i - \alpha - s_i\beta - z_i\gamma) \quad (8)$$

If β is given, then the estimates of α and γ can be obtained using traditional quantile regression parameter estimation methods, that is,

$$(\hat{\alpha}(\beta, \tau) \hat{\gamma}(\beta, \tau)) := \arg \min_{\alpha, \gamma} Q_n(\tau, \alpha, \beta, \gamma) \quad (9)$$

The method of estimating $\beta(\tau)$ is to make the value of $\hat{\gamma}(\beta, \tau)$ closest to 0, which is expressed by the formula:

$$\hat{\beta}(\tau) = \arg \inf_{\beta} W_n(\beta) \quad (10)$$

Where $W_n(\alpha) := n[\hat{\gamma}(\beta, \tau)'] \Lambda(\beta) [\hat{\gamma}(\beta, \tau)]$, and $\Lambda(\beta)$ is a positive definite matrix. Here, for simplicity of calculation, let it be identity matrix.

5. Empirical Research

5.1 Sample selection and data sources

Collecting social media data is an important step to verify the relationship between different users' sentiment and the stock market. Using Weibo's API, we obtain all Weibo content containing the keywords of "stock market" in Weibo published from March 1, 2018 to October 31, 2018, as well as the basic information of the publishers. In this experiment, the sentiment analysis module of ROST Content Mining (ROST CM for short) [22] is used to analyze the sentiment of daily Weibo. ROST CM's content mining system has strong support for Chinese language and is rich in functions, providing great technical support for Chinese-based data mining and knowledge discovery [7]. ROST CM scores each Weibo, and divides the emotional tendency into seven levels: general

positive sentiment (0 ~ 10), moderate positive sentiment (10-20), highly positive sentiment (more than 20), neutral sentiment (0), general negative sentiment (-10 ~ 0), moderate negative sentiment (-20 ~ -10), and highly negative sentiment (below -20).

At the same time, the daily closing price data of the Shanghai Composite Index from March 1, 2018 to October 31, 2018 are selected. The required data are obtained from the wind database. A total of about 192,000 original posts are collected in this study. Weibo users are classified into authenticated users and non-authenticated users based on whether the user has authentic verification. The Weibo sentiments of the two types of users are matched with stock data. Finally, 161 sets of daily data are obtained.

In order to gain a preliminary and intuitive understanding of the daily Weibo sentiment distribution, Weibo sentiments for two days are selected randomly to plot intra-day histograms of Weibo sentiments for two types of users (See Figure 1). The sentiment classification is chosen according to the criterial of ROST CM. It can be seen from the figure that the two types of users both posted more positive posts than negative ones in these two days. In addition, most negative sentiment values are concentrated between -10 and 0, that is, general negative sentiment, and most positive sentiment are concentrated in the range of 0 to 10 and greater than 20, that is, general positive emotions and highly positive emotions. This shows that in general, the sentiment in Weibo about the stock market in the day is mainly positive, with the negative sentiment being mostly general and the positive sentiment being mostly general and highly. It can also be seen that the total number of authenticated users is less than that of non-authenticated users. And non-authenticated users account for more extreme sentiment (ie, highly negative and highly positive sentiment) than authenticated user

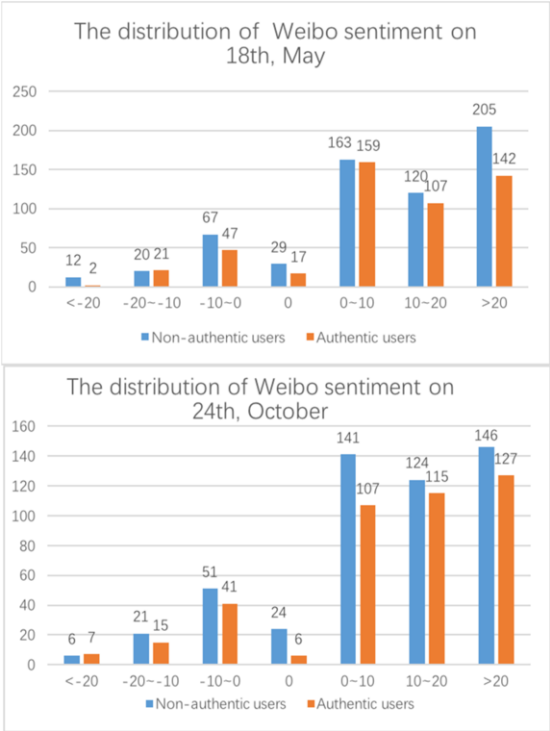


Figure 1: Intra-day histograms of Weibo sentiments

5.2 Variable statistical description

In previous studies on the relationship between Internet sentiment and the stock market, stock market returns are typically used as explanatory variables, and sentiment values or sentiment fluctuations are used as explanatory variables. The Weibo sentiment and sentiment fluctuations of non-authenticated users and authenticated users are marked as SU, SV, DSU, DSV, and the stock market return of the Shanghai Composite Index is marked as R, where:

$$R_t = \ln(P_t) - \ln(P_{t-1}) \quad (11)$$

$$DSU_t = SU_t - SU_{t-1} \quad (12)$$

$$DSV_t = SV_t - SV_{t-1} \quad (13)$$

Descriptive statistics of each variable are shown in Table 1.

Table 1: Descriptive statistics of variables

Variable	Notation	Mean	Median	Maximum	Minimum	Standard deviation	Skewness	Kurtosis
<i>Stock market return</i>	R	-0.0006	-0.0008	0.0174	-0.0233	0.0056	-0.3847	4.9488
<i>Sentiment of non-authentic users</i>	SU	10.9937	10.8345	22.6342	-9.2206	3.1955	-1.7020	15.8762
<i>Sentiment of authentic users</i>	SV	11.1521	11.2422	16.7606	1.0640	2.2543	-0.5843	4.5054
<i>Sentiment fluctuation of non-authentic users</i>	DSU	0.0429	-0.0574	18.8107	-16.3306	3.9400	0.1306	9.3041
<i>Sentiment fluctuation of authentic users</i>	DSV	0.0184	-0.1245	8.7466	-12.8903	2.9723	-0.4157	4.6431

It can be seen from the table that the mean and median sentiment scores of authenticated users on Weibo are higher than those of non-authenticated users, which indicates that the sentiment of authenticated users on Weibo is more positive than that of non-authenticated users. Sentiment of non-authenticated users are more extreme and discrete than that of authenticated users. It can also be seen from the sentiment fluctuations that the sentiment fluctuations of non-authenticated users are greater than that of authenticated users, which presents a more dispersed trend. In short, from the descriptive statistics of the variables, we can initially and intuitively see that the sentiment of non-authenticated users is more negative and extreme and the sentiment fluctuation is more volatile. On the other hand, the sentiment of authenticated users is more positive and concentrated, and the sentiment fluctuation is more stable.

In order to have a more intuitive understanding of each variable, a line chart of stock market returns and sentiment of two types of users is drawn (Figure 2 and Figure 3 represent the relationship between stock market returns and sentiment and sentiment fluctuation respectively). It can be seen from the figures that there is some correlation between the sentiment and sentiment fluctuation of the two types of users, and that the sentiment fluctuation of non-authenticated users is greater and the sentiment value is more extreme, while the sentiment of authenticated users is more stable and neutral. At

the same time, the images also intuitively show that there is a certain correlation between stock market returns and sentiment of the two types of users, which requires further exploration with econometric models.



Figure 2: Time series of stock market returns and the sentiment

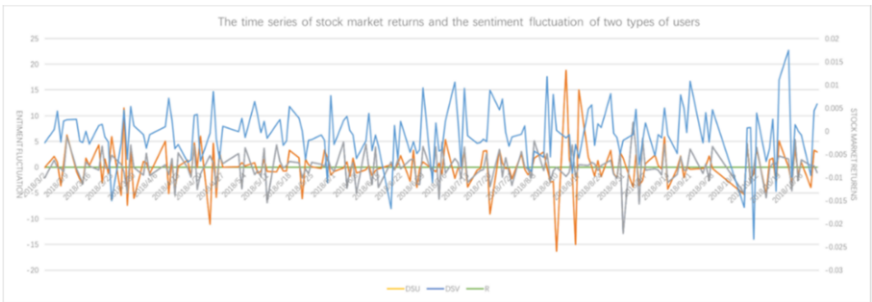


Figure 3: Time series of stock market returns and the sentiment fluctuation

5.3 Evaluation Results

5.3.1 Unit root test

In the application of classical linear regression in time series, in order to avoid the problem of false regression or pseudo regression, the stability of the time series needs to be considered first. Here, the ADF test is used, and the lag order is determined by the SIC criterion. The results of the unit root test are shown in Table 2. Each series has a P value of 0.0000, indicating that each series is a stationary series.

Table 2: Results of unit root test

Variable	Test form	ADF	P-value	Conclusion
R	(C, T, 0)	-14.3375	0.0000	Stationary
SU	(C, T, 0)	-4.8551	0.0001	Stationary
SV	(C, T, 2)	-4.9931	0.0000	Stationary
DSU	(C, T, 0)	-17.4123	0.0000	Stationary
DSV	(C, T, 1)	-16.2435	0.0000	Stationary

Note: C, T, and K in the test form represent the constant term, the time trend term, and the lag order, respectively

5.3.2 Results of quantile regression

In order to characterize the impact of Weibo sentiment and sentiment fluctuation of two types of users on stock market returns at different quantiles, the following two quantile regression models are established, with stock market returns as the explained variables and Weibo sentiment and sentiment fluctuation as explanatory variables respectively. In order to examine whether the impact of sentiment on stock market returns is transient or continuous, the current and previous period values of sentiment or sentiment fluctuation are included in the model. The specific model is as follows:

$$R_t = \alpha_0(\tau) + \alpha_1(\tau)R_{t-1} + \alpha_2(\tau)SU_t + \alpha_3(\tau)SU_{t-1} + \alpha_4(\tau)SV_t + \alpha_5(\tau)SV_{t-1} + \mu(\tau)_t \quad (14)$$

$$R_t = \beta_0(\tau) + \beta_1(\tau)R_{t-1} + \beta_2(\tau)DSU_t + \beta_3(\tau)DSU_{t-1} + \beta_4(\tau)DSV_t + \beta_5(\tau)DSV_{t-1} + \varepsilon(\tau)_t \quad (15)$$

Where τ is the quantile, $\alpha_0(\tau)$ and $\beta_0(\tau)$ are the intercept terms, $\alpha_1(\tau)$ and $\beta_1(\tau)$ are the effects of the stock market returns of the previous period on the current stock market returns, $\alpha_i(\tau)$ and $\beta_i(\tau)$ ($i = 2, 3, 4, 5$) are the regression coefficients corresponding to the explanatory variables at different quantiles of stock market returns, and $\mu(\tau)_t$ and $\varepsilon(\tau)_t$ are random error terms. This paper selects [0.1, 0.9] as the study interval, and takes a quantile every 10%. The results obtained by the two models are shown in the Table3. For comparative analysis, the results of the least squares estimation are also listed.

Table 3: Results of quantile regression

Panel1					
	R(-1)	SU	SV	SU(-1)	SV(-1)
OLS	-.07982 (.00301)	.00023 (.00014)	.00060*** (.00020)	.00006 (.00015)	.00060 (.00021)
0.1	-.16987* (.16797)	.00021 (.00040)	.00044 (.00054)	.00007 (.00020)	-.00019 (.00046)
0.2	-.15343 (.12804)	.00018 (.00031)	.00068* (.00036)	.00013 (.00017)	-.00001 (.00032)
0.3	-.12618 (.11271)	.00033 (.00036)	.00048 (.00034)	.00008 (.00023)	-.00020 (.00022)
0.4	-.10300 (.09629)	.00028 (.00023)	.00047* (.00027)	.00004 (.00017)	-.00014 (.00019)
0.5	-.06470 (.08900)	.00026 (.00019)	.00071** (.00029)	.00013 (.00013)	-.00018 (.00018)
0.6	-.10764 (.08006)	.00031** (.00015)	.00047*** (.00017)	.00034 (.00012)	-.00017 (.00018)
0.7	-.05769 (.08115)	.00030** (.00014)	.00048*** (.00016)	-.00033 (.00015)	-.00005 (.00017)
0.8	.01151 (.07758)	.00026** (.00012)	.00058*** (.00017)	-.00011 (.00029)	.00005 (.00020)
0.9	-.14617 (.17622)	.00017 (.00015)	.00083*** (.00019)	.00027** (.00012)	-.00040 (.00046)

Panel2					
	R(-1)	DSU	DSV	DSU(-1)	DSV(-1)
OLS	-.07228 (.07988)	.00017 (.00012)	.00068*** (.00018)	.00017 (.00012)	.00050*** (.00017)
0.1	-.11930 (.17562)	.00012 (.00043)	.00021 (.00043)	.00011 (.00040)	-.00001 (.00041)
0.2	-.16431* (.08979)	.00006 (.00011)	.00064** (.00030)	.00021** (.00009)	.00037 (.00026)
0.3	-.13211* (.07036)	.00017 (.00011)	.00071*** (.00015)	.00014 (.00011)	.00049*** (.00014)
0.4	-.11697 (.07574)	.00020* (.00011)	.00058*** (.00016)	.00017 (.00010)	.00047*** (.00012)
0.5	-.08531 (.07447)	.00022* (.00012)	.00054*** (.00015)	.00021 (.00014)	.00040*** (.00012)
0.6	-.05060 (.07075)	.00027** (.00012)	.00049*** (.00013)	.00023 (.00015)	.00036*** (.00012)
0.7	-.00142 (.06837)	.00031*** (.00011)	.00049*** (.00013)	.00018 (.00015)	.00037*** (.00012)
0.8	-.05644 (.09029)	.00021 (.00015)	.00085** (.00034)	.00007 (.00015)	.00071** (.00032)
0.9	-.03963 (.12648)	.00023 (.00016)	.00124*** (.00034)	.00030 (.00024)	.00078* (.00044)

Note: *** indicates that the statistic is significant at the 1% level, ** indicates that the statistic is significant at the 5% level, * indicates that the statistic is significant at the 10% level, and the standard error is shown in parentheses (the same below).

Table 4: Results of IVQR model

Panel1						
	R(-1)	SU	SV	R(-1)	DSU	DSV
IV	-.12812* (.07576)	.00022 (.00014)	.00062*** (.00020)	-.12812* (.07523)	.00017 (.00012)	.00067*** (.00018)
0.1	-.23537 (.22715)	.00025 (.00043)	.00034 (.00061)	-.19643 (.22577)	-0.00001 (.00037)	.00009 (.00053)
0.2	-.17859 (.11510)	.00010 (.00022)	.00058* (.00031)	-.13451 (.11851)	-0.00001 (.00020)	.00060** (.00028)
0.3	-.12901 (.08517)	.00031* (.00016)	.00055** (.00023)	-.16611* (.08295)	.00018 (.00014)	.00066*** (.00020)
0.4	-.15312* (.07797)	.00027* (.00015)	.00046** (.00021)	-.14413** (.06657)	.00014 (.00011)	.00068*** (.00016)
0.5	-.12004* (.07000)	.00016 (.00013)	.00081*** (.00019)	-.09345 (.06143)	.00017* (.00010)	.00064*** (.00015)
0.6	-.13921* (.07444)	.00027* (.00014)	.00059*** (.00020)	-.10802 (.06747)	.00024** (.00011)	.00061*** (.00015)
0.7	-.09560 (.09428)	.00029* (.00018)	.00050* (.00025)	-.08008 (.08932)	.00028* (.00015)	.00054** (.00021)
0.8	-.04683 (.11993)	.00035 (.00023)	.00056* (.00032)	-.01329 (.12287)	.00027 (.00020)	.00079*** (.00029)
0.9	-.15867 (.15327)	.00023 (.00029)	.00080* (.00041)	-.23790* (.14322)	.00023 (.00024)	.00079** (.00034)

OLS regression results show that the sentiment and sentiment fluctuation of both types of users are positively correlated with stock market returns, but only the sentiment and sentiment fluctuation of authenticated users have significant correlations with stock market returns, while those of non-authenticated users and stock market returns are not significant. Obviously, OLS regression focuses on discussing the average effect of independent variables on the dependent variable. In the application of practical problems, investors and decision makers often pay more attention to the extreme stock market conditions, and it is difficult for the mean regression method to effectively capture these characteristics to obtain incomplete statistical inference results [23]. In addition, the relationship between Weibo sentiment and stock market returns may be manifested as a part of the quantile interval that is positively significantly correlated, partly negatively correlated, and partly non-significantly correlated, so that the positive and negative correlations cancel each other out. As a result, the relationship between Weibo sentiment and stock market returns in OLS returns is not significant. Therefore, a quantile regression model is needed to explore the heterogeneity of the relationship between Weibo sentiment and stock market returns at different quantiles in order to obtain more accurate findings.

In the results of quantile regression, there exist significant positive quantiles in the relationship between the sentiment and stock market returns of both types of users, but the significant quantiles are different. For non-authenticated users, the significant quantile intervals of current sentiment are concentrated in $[0.6, 0.8]$, and that of the sentiment in the previous period exist only in a few quantiles (0.9). For authenticated users, significant positive correlation between current sentiment and stock market returns are distributed at various quantiles, but there is no significant correlation between sentiment in the previous period and stock market returns in the selected quantiles. It can be concluded that for non-authenticated users, only the relatively moderate positive sentiment in the current period is significantly and positively correlated with stock market returns, while for authenticated users, both negative and positive sentiment in the current period have significant positive correlations with stock market returns. However, for both types of users, the sentiment in the previous period is almost irrelevant with stock market returns.

In addition, there exist significant positive quantiles in the relationship between the sentiment fluctuation and stock market returns of both types of users. For non-authenticated users, the significant correlation between current sentiment fluctuation and stock market returns is concentrated in the middle quantiles (0.4-0.7), and the sentiment fluctuation in the previous period is not related to stock market returns. For authenticated users, the significant correlation between the sentiment fluctuation in current and previous period and stock market returns is distributed at various quantiles. It indicates that for non-authenticated users, only the moderate current sentiment fluctuation is significantly and positively correlated with stock market returns, while sentiment fluctuation on lag 1 is not correlated with stock market returns. For authenticated users, the sentiment fluctuations in current and previous period are significantly and positively correlated with stock market returns in most cases.

5.3.2 Results of instrumental variable quantile regression

Since the sentiment of the day can affect stock market returns, stock market returns can also affect the sentiment of the day. To be specific, the higher the quantile level of stock market returns, the more positive the sentiment. As a result, the quantile regression model

is endogenous. The slope parameter obtained can only reflect the correlation between sentiment or sentiment fluctuation and stock market returns at different quantiles. It cannot objectively and accurately measure the impact of sentiment or sentiment fluctuation on stock market returns. In order to eliminate the endogenous nature, the IVQR model is constructed by using the sentiment or sentiment fluctuation of the previous period as an instrumental variable. The results obtained are shown in the Table 4. For comparison, the results of a common instrumental variable model are also listed.

It can be seen from Table 4 that with the increase in sentiment and sentiment fluctuation, stock market returns will increase, but on the average, only the sentiment and sentiment fluctuation of authenticated users have a significant and positive impact on stock market returns. From the results of IVQR model, the significant positive impact of the sentiment and sentiment fluctuation of the authenticated users on stock market returns is distributed at various quantiles, while the significant effects of non-authenticated users have certain conditions. For example, for non-authenticated users, the significant impact of sentiment on stock market returns is concentrated in the middle quantile (0.3-0.7), indicating that more neutral sentiments have a significant positive impact on stock market returns, while more extreme sentiments have no significant effect on stock market returns. The significant impact of non-authenticated users' sentiment fluctuations on stock market returns is also concentrated in the middle quantile (0.5-0.7), indicating that only moderate sentiment fluctuations have a significant impact on stock market returns, while large or small ones have no significant effects. In addition, the coefficients of sentiment and sentiment fluctuation of authenticated users at each quantile is larger with higher significance than that of non-authenticated users with the same quantile, indicating that the sentiment and sentiment fluctuation of authenticated users have a stronger impact on stock market returns than those of non-authenticated users.

6. Discussion

Through the analysis and summary of the above results, we can draw the following main conclusions.

(1) Generally speaking, both sentiment and sentiment fluctuation have a positive effect on stock market returns, which is the same as the conclusions drawn by most previous studies [7,14]. When the user's sentiment on Weibo is optimistic (pessimistic), the stock market returns on that day will rise (fall) in a short period of time, and this effect is short-lived. The stock market returns on the next day are not affected by the Weibo sentiment in the previous day. The reason may be that sentiment acts as a kind of information, it takes a short time to integrate the stock market price in social media, so the Weibo sentiment can only affect the stock market performance of the current day, and the stock market performance of the next day is affected by the Weibo of the second day. According to previous research, the greater the sentiment fluctuation of social media users, the greater the stock market trading volume [14], and the larger the trading volume, the higher the stock market returns [24], indicating that the sentiment fluctuation may affect stock market returns through transaction volume. Since the social media sentiment fluctuation of the day can significantly affect the trading volume on the second day [14], the impact of sentiment fluctuation on stock market returns will continue to the next day, which also explains the tendency to inertial effect in Chinese stock market returns to some degree.

(2) Compared with non-authenticated users, the sentiment and sentiment fluctuation of authenticated users have a stronger and longer-term impact on stock market returns. From the above models, the coefficients of the sentiment and sentiment fluctuation of the authenticated users are significantly higher than those of the non-authenticated users. In addition, the sentiment fluctuation of authenticated users can also have a positive impact on the stock market returns in the next period, indicating that authenticated users have a longer-term impact on stock market returns. This may be due to the nature and characteristics of the authenticated user. Compared with ordinary users, authenticated users are more concerned and more authoritative. The proportion of high-impact users is larger and they have stronger influence [25]. Because the authenticated users have a certain popularity, the information they disseminate is considered valuable by followers, and it is easy to cause onlookers and reposts [26]. Therefore, the sentiment fluctuation of authenticated users can be spread by followers, which leads to a longer-term impact on stock market changes. In contrast, non-authenticated users have limited influence and popularity, and their emotions cannot be spread to a large extent like authenticated users, so their emotional changes cannot affect future stock market returns.

(3) Only relatively moderate sentiment and sentiment fluctuation of non-authenticated users have a positive impact on stock market returns, while authenticated users have fewer restrictions. The significant impact of non-authenticated users' sentiment on stock market returns is concentrated in the middle quantile, indicating that more moderate sentiment has an impact on stock market returns, while more extreme emotions cannot affect stock market returns. The reason may be that different sentiment of different users have different transmission effects [27]. Chua and Banerjee [28] have pointed out that too much emotional expression in the review information makes the information audience doubt its motivation for communication and underestimate its usefulness. Therefore, in the process of disseminating information about the stock market on Weibo, for ordinary non-authenticated users with low credibility, when the contents they post have excessive extreme sentiment and excessive sentiment fluctuation, the information audience may doubt the authenticity of the content, and feel that the publisher of the information is just venting his emotions and not passing on the facts. Therefore, the audience's willingness to forward and disseminate such content is relatively low, and such content cannot affect stock market returns. Authenticated users generally have high professionalism and popularity. When they release extreme sentiment, or have large sentiment fluctuation, the audience will give higher trust, and the emotions they express are more likely to infect other people [29,30]. Therefore, extreme sentiment and large sentiment fluctuations of authenticated users can also affect stock market returns.

In addition, it is found that at the low quantiles of stock market returns, the sentiment and sentiment fluctuation of non-authentic users have no significant effect on stock market returns. According to previous research, at low quantiles of market returns, the market is dominated by rational investors, while irrational traders take a wait-and-see attitude [31]. According to the analysis of the previous descriptive statistical variables, the sentiment of non-authenticated users is more extreme, and the sentiment fluctuation is also larger, which can be considered as irrational traders. The sentiment of authenticated users is more neutral and stable, with less sentiment fluctuation. In addition, many financially authenticated users have more financial knowledge and market information, which can be considered as rational traders. Therefore, at the low quantiles of stock market returns, the market is dominated by rationally authenticated users, and only the sentiment and sentiment fluctuation of authenticated users can affect stock

market returns. At the middle and higher quantiles of stock market returns, both authenticated and non-authenticated users participate in the market. The sentiment and sentiment fluctuation of both types of users are driving further changes in stock returns. However, when the sentiment fluctuation of non-authenticated users is greater than a certain level, credibility is suspected and the level of transmission is reduced. As a result, non-authenticated users can not have a significant impact on stock market returns.

7. Conclusion

The purpose of this study is to explore the impact of sentiment and sentiment fluctuation of different users in social media on stock market performance. The study takes Weibo and the Shanghai Composite Index as the research object, divides Weibo users into two categories: authenticated users and non-authenticated users, and calculates the sentiment and sentiment fluctuation of the two types of users. Then, it uses quantile regression and IVQR model to analyze the impact of sentiment and sentiment fluctuation of the two types of users on stock market returns. The research results show that Weibo sentiment and sentiment fluctuation have a positive impact on stock market returns, but the effects of the two types of users are different. Due to the authority and popularity of authentic users, their sentiment has a stronger and longer-term impact on stock market returns. Due to the limited influence and credibility of non-authenticated users, only moderate sentiment and sentiment fluctuation have a positive impact on stock market returns, while extreme sentiment and large sentiment fluctuations cannot impact stock market returns significantly.

The significance of this study is mainly reflected in two aspects. The first is to provide new evidence for the relationship between social media sentiment and sentiment fluctuation and stock market performance. The results of the quantile regression of instrumental variables show that whether they are authenticated users or ordinary users in social media, their sentiment and sentiment fluctuation can help increase their understanding of stock market performance. By including them in the prediction model, the interpretation ability and prediction accuracy of the model can be improved. In addition, it can also be seen that different categories of users in social media have different effects on the stock market. The impact of authentic users on stock market performance is stronger and longer-term, while that of non-authentic users has certain conditions, that is, only moderate sentiment and sentiment fluctuations can affect stock market performance. By distinguishing the different impacts of different users on stock market returns, it helps to further understand the mechanism of social media sentiment, and it also has certain practical significance for social media platforms and users. For example, when big event occurs, the public opinion orientation of authenticated users should be focused on control, and in particular, they should pay attention to the information that they release with extreme sentiment and large sentiment fluctuation. For non-authenticated users, in order to have a wider range of information, they should avoid revealing excessively extreme sentiment and obvious sentiment fluctuation, thereby increasing the credibility of the information.

Although this study has made some meaningful findings based on new perspectives and new measurement methods, there are still shortcomings. The first is using sentiment analysis software to directly analyze Weibo content. The accuracy rate is still insufficient, so there is a certain error in the calculated sentiment value. Further research can improve the accuracy of sentiment calculations by manually labeling a large number of corpora

and using machine learning methods. The second is considering only two categories of authenticated and non-authenticated users. In fact, there are many types of user categories in social media. Subsequent research can further segment social media users and explore the impact of more different user emotions on stock market performance.

References

- [1] Horne JC Van, Parker GGC. The Random-Walk Theory: An Empirical Test. *Financ Anal J.* 1967;23(6).
- [2] Schmeling M. Investor sentiment and stock returns: Some international evidence. *J Empir Financ.* 2009;16(3).
- [3] Chan SY, Fong WM. Individual investors' sentiment and temporary stock price pressure. *J Bus Financ Account.* 2004;31(5–6).
- [4] Baker M, Wurgler J. Investor sentiment and the cross-section of stock returns. *J Finance.* 2006;61(4).
- [5] Bathia D, Bredin D. An examination of investor sentiment effect on G7 stock market returns. *Eur J Financ.* 2013;19(9).
- [6] Zhang X, Fuehres H, Gloor PA. Predicting Stock Market Indicators Through Twitter “I hope it is not as bad as I fear.” *Procedia - Soc Behav Sci.* 2011;26.
- [7] Huang R, Zuo W, Bi L. Stock market prediction based on Weibo sentiment information. *J Manag Eng.* 2015;29(1):47–52.
- [8] Xuemei T, Shengqiang L. Research on the Influence of Emotional Information on the Dissemination of Public Opinion Events. *Inf Mag.* 2018;37(12):128–33.
- [9] Lin D. Study on the Dynamic Relationship between Shanghai Stock Market Returns and Trading Volumes: An Analysis Based on Quantile Regression Models. *Explor Financ Theory.* 2014;157(5):48–52.
- [10] Chernozhukov V, Hansen C, Jansson M. Inference approaches for instrumental variable quantile regression. *Econ Lett.* 2007;95(2).
- [11] Renault T. Intraday online investor sentiment and return patterns in the U.S. stock market. *J Bank Financ.* 2017;84.
- [12] Zhou S, Shi X, Sun Y, Qu W, Shi Y. Stock Market Prediction Using Heat of Related Keywords on Micro Blog. *J Softw Eng Appl.* 2013;06(03).
- [13] Cheng Y, Lin J. Investors' Ups and Downs and Social Market Index in Social Media. *Manage Sci.* 2013;(5):111–9.
- [14] Siganos A, Vagenas-Nanos E, Verwijmeren P. Divergence of sentiment and stock market trading. *J Bank Financ.* 2017;78.
- [15] See-To EWK, Yang Y. Market sentiment dispersion and its effects on stock return and volatility. *Electron Mark.* 2017;27(3).
- [16] Fama EF. Efficient Capital Markets: A Review of Theory and Empirical Work. *J Finance.* 1970;25(2).
- [17] Zhou Z. Behavioral Finance-Theory and Application. 2004.
- [18] Shefrin H. Beyond Greed and Fear. *Beyond Greed and Fear.* 2002.
- [19] Gao Q. On the Impact of Investor Sentiment on the Stock Market. *Econ Asp.* 2005;(4):34–6.
- [20] Xu Q, Bo Z, Jiang C, Xu Q, Bo Z, Jiang C. Research on the Relationship between Network Emotion and Stock Market Returns Based on Quantile Granger Causality. *Manage Sci.* 2017;30(3):147–60.
- [21] Koenker R, Bassett G. Regression Quantiles. *Econometrica.* 1978;46(1):33.
- [22] Shen Y. ROST content mining system: software for content mining and analysis. 2008.
- [23] Chen X, Huang Y. Research on the linkage relationship between the international crude oil market and the stock market: empirical evidence based on quantile regression. *Financ Theory Pract.* 2017;38(5):53–8.
- [24] Ren Y, Li X. A Study on the Dynamic Relationship between Return and Trading Volume in China's Stock Market—A Quantile Regression (IVQR) Model Based on Instrumental Variables. *China Manag Sci.* 2017;25(8):11–8.
- [25] Yu J. Empirical analysis of user influence in the process of Weibo communication. *J Inf.* 2013;(8):57–61.
- [26] Wang P, Xie Y. An Empirical Study of Weibo Opinion Leaders in Public Emergencies: A Case Study of the Wenzhou Motor Vehicle Accident. *Mod Commun (Journal Commun Univ China).* 2012;34(3):82–8.
- [27] Tang X, Lai S. Research on the Influence of Emotional Information on the Spread of Public Opinion Events. *J Inf.* 2018;37(12):128–33.

- [28] Chua AYK, Banerjee S. Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. Vol. 66, *Journal of the Association for Information Science and Technology*. 2015. p. 354–62.
- [29] Jian P, Yi L, Hong L, Yingjie L, Weitong W. Emotional leadership: concept, measurement, development and model construction. *Adv Psychol Sci*. 2014;22(11):1757–69.
- [30] Ping Q, Ying Z, Ming T, Yong Z, Tong C. Research on the Influence of Mobilization Methods of Internet Opinion Leaders on Internet Cluster Behavior Participation——Based on the Experimental Research under the Background of Product Harm Crisis. *Manage World*. 2016;(7):67–79.
- [31] Dong X, Zhang X, Liu W. The Interactive Relationship between Investor Sentiment and Stock Market Returns: A Study Based on Quantile Regression. *Econ Manag*. 2013;(6):103–11.