Received: 28 September 2020 | Revised: 11 January 2021 | Accepted: 11 April 2021

DOI: 10.1002/for.2777

RESEARCH ARTICLE

WILEY

Predicting stock market volatility based on textual sentiment: A nonlinear analysis

Weiguo Zhang[1] | **Xue Gong**[1] | **Chao Wang**[1] | **Xin Ye**[2]

[1]School of Business Administration, South China University of Technology, Guangzhou, China

[2]School of Mathematics and Statistics, Wuhan University, Wuhan, China

Correspondence
Xue Gong, School of Business Administration, South China University of Technology, Guangzhou, China.
Email: fixgong@outlook.com

Funding information
Foundation for Key Program of Ministry of Science and Technology of China, Grant/Award Number: 2020AAA0108404; National Natural Science Foundation of China, Grant/Award Numbers: 71720107002, 71771091; Financial Service Innovation and Risk Management Research Base of Guangzhou

Abstract

This paper investigates whether and how investor sentiment affects stock market volatility forecasting from a nonlinear theory perspective. With the use of a novel dataset that contains massive articles about stock market analysis obtained from a Chinese investors' community, we construct four sentiment indices to measure investor sentiment by applying textual analysis techniques. Differing from the developed market, we find that the investor sentiment from an emerging market causes stock volatility by a nonlinear pattern rather than a linear style. Furthermore, we show that the investor sentiment improves stock volatility prediction based on the long short-term memory model. And the predictability is still significant after considering another sentiment proxy variable. Finally, we demonstrate that this improvement of predictive performance is meaningful from an economic point of view.

KEYWORDS

investors' community, machine learning, textual (investor) sentiment, volatility prediction

1 | INTRODUCTION AND LITERATURE REVIEW

May what the media say have an impact on the stock market? Some recent facts may answer the question. On November 22, 2017, the stock price of listed company Red-Yellow-Blue Kindergarten (the RYB Education) was steep fall due to the exposure of children abuse events. Coincidentally, On May 6, 2019, the president of the United States Trump announced an additional 15% tariff on China through Twitter, followed by a 6% plunging on A-shares (the Shanghai–Shenzhen stock market in China). On December 13, 2019, China–U.S. agreed to a limited deal to halt the trade war, followed by a 2% rising on A-shares. These news tell the truth that internet messages are able to influence financial market activities. Under the same event, that is, China–U.S. trade war, the market acts differently on bad news and good news (6% plunging for disagreement vs. 2% rising for agreement). Previous works have provided some evidence that information extracted from textual sources, such as news, investor forum, and social network, can affect stock market movement (see, for example, Bollen et al., 2011; Yu et al., 2013; Huang & Liu, 2020).

As reported by the Xinhua news agency in March 2020, the number of individual investors on the A-share market is close to 160 million. According to the behavioral finance theory, for instance herd behavior, individual investors are more inclined to rely on information from the internet network for decision making, which has an impact on stock market activity. In this paper, we concentrate on the relationship between investor sentiment (measured by textual sentiment) and future stock market volatility. As an important part of behavioral finance theory, investor sentiment can be defined as investors' false beliefs about an asset value (Zhou, 2018). We provide a concise and typical literature summary about investor sentiment in financial market as shown in Table 1. Three category measures of investor sentiment can be summarized as follows: (i) Survey index, based on

TABLE 1 Related work about investor sentiment

Literature	Sentiment: Method or representation	Frequency	Source	Type	Market	Period	Model	Value
Lee et al. (2002)	II index	Weekly	DRI	Survey	US	01/05/1973–10/06/1995	Modified GARCH	Volatility & excess return
Brown and Cliff (2004)	AAII, II, et al.	Monthly & Weekly	Do not mention	Survey	US	03/1965–12/1998 (M) 07/24/1987–12/18/1998 (W)	VAR	Return
Antweiler and Frank (2004)	Aggregation function	15-minute	Yahoo! Finance & Raging Bull	Text	US	Full year 2000	GARCH-class & regression	Return, volatility, & trading volume
Baker and Wurgler (2006)	PCA proxy	Monthly	CRSP	Market data	US	1963–2001	Regression	Return
Das and Chen (2007)	Six classifiers[a]	Daily	Yahoo! Finance	Text	US	07/2001–08/2001	Regression	Volatility & trading volume
Bollen et al. (2011)	OF and GPMS	Daily	Twitter	Text	US	02/28/2008–12/19/2008	SOFNN	Closing values
Yu et al. (2013)	Following AF[b]	Daily	Forum, news, microblog, Newspapers, television & business magazines	Text	US	07/01/2011–09/30/2011	Regression	Return & risk
Huang et al. (2015)	PLS proxy	Monthly	JeffreyWurgler's website	Market data	US	07/1965–12/2010	Regression	Return
Nguyen et al. (2015)	Human sentiment, LDA-based method, JST-based method & Aspect-based sentiment	Daily	Yahoo! Finance	Text	US	07/23/2012–07/19/2013	SVM	Adjusted close price

TABLE 1 (Continued)

Literature	Sentiment: Method or representation	Frequency	Source	Type	Market	Period	Model	Value
Johnman et al. (2018)	Dictionary-based	Daily	Guardian Media Group	Text	UK	01/01/2000–01/06/2016	Regression	Excess return & volatility
Zhang et al. (2018)	SnowNLP & NB algorithm	Daily	Xueqiu	Text	China	12/2014–05/2015	SVM & MLP	Price movements
Behrendt and Schmidt (2018)	From Bloomberg	5-minute	Twitter	Text	US	06/18/2015–12/29/2017	VAR & HAR	Absolute return
Ren et al. (2018)	Following others	Daily	Sina and Eastmoney	Text	China	06/17/2014–06/07/2016	SVM	Stock market movement
Jiang et al. (2019)	Dictionary-based	Monthly	10-Ks, 10-Qs & conference call	Text	US	01/2003–12/2014	Regression	Return
Zoen Git Hiew et al. (2019)	BERT method	Daily	Weibo	Text	China	2016–2018	VAR & LSTM	Return
Audrino et al. (2020)	Deep-MLSA	Daily	Twitter and StockTwits	Text	US	2012–2016	HAR	Stock volatility

Note: This table provides a concise and typical literature summary about investor sentiment in the financial field. We use some abbreviations as described in the following due to space limitations here. Abbreviations (sort from upper left to low right): II, Investors Intelligence; AAII, American Association of Individual Investors; PCA, Principal Component Analysis; OF, OpinionFinder; GPMS, Google-Profile of Mood States; PLS, Partial Least Squares; LDA, Latent Dirichlet Allocation; JST, Joint Sentiment/Topic; NB, Naive Bayes; BERT, Bidirectional Encoder Representations from Transformers GARCH, Generalized Autoregressive Conditional Heteroskedasticity; VAR, Vector autoregression; SOFNN, Self-Organizing Fuzzy Neural Network; SVM, Support Vector Machine; MLP, multilayer perceptron; HAR, Heterogeneous Autoregression; LSTM, Long–Short-Term Memory.

[a]Six classifiers refer to Naive Classifier, Vector Distance Classifier, Discriminant-Based Classifier, Adjective-Adverb Phrase Classifier, Bayesian Classifier, and Voting Amongst Classifiers.

[b]That follows Antweiler and Frank (2004).

market survey. It is a direct sentiment measure released by professional institution, such as the American Association of Individual Investors (i.e., AAII)[1] and Investors Intelligence (i.e., II) (see Lee et al., 2002; Wang, 2018, and among others). (ii) Proxy index, constructed by market data; for example, Baker and Wurgler (2006) (BW) selected six proxies with their lags and then utilized the first component from the result of principal component analysis (PCA) to measure investor sentiment (BW index below).[2] The BW index is cited extensively by scholars (see, for example, Yu & Yuan, 2011; Stambaugh et al., 2012; Huang et al., 2015). In contrast to the survey index, the proxy index is an indirect sentiment measurement that uses market indicators to reflect investor sentiment. (iii) Textual sentiment index, obtained through textual analysis; it has a higher frequency (daily or intraday) relative to the survey index and proxy index, which are usually at weekly or monthly frequency. The effectiveness of textual sentiment index relies on the selection of text source and sentiment analysis method. For text source, a recent survey concluded by Kearney and Liu (2014) presents that the public corporate disclosure, internet messages, and media articles are corpus sources of textual analysis in the financial field. For textual sentiment analysis, determining text polarity, that is, sentiment classification, such as positive and negative sentiments, and transforming text data into a numerical format, that is, text conversion, are two key steps.

Machine learning (ML) techniques and dictionary-based methods are well-known in textual sentiment analysis. The former is divided into three steps: (i) preparing training and test dataset; (ii) choosing a classifier model and training it on training data; and (iii) classifying new text by using the well-trained classifier. This method, to some degree, is difficult to explain and repeat because of black-box operation. Another, the latter focuses on classifying each word or phrase in sentences or documents into a corresponding sentiment category (positive, negative, and/or others) based on a prepared sentiment dictionary. By reviewing existing sentiment dictionaries of both English and Chinese language, we find that the sentiment dictionary of financial domain in English is developed better than that in Chinese. For example, Loughran and McDonald (2016) concluded four kinds of popular English sentiment dictionaries: Henry (2008), Harvard GI, Diction, and L&M dictionary. Among these word lists, of particular note is the L&M dictionary proposed by Loughran and McDonald (2011). This dictionary is constructed by using a corpus extracted from corporate disclosure, including 354 positive words and 2355 negative words. While for the Chinese sentiment dictionary, there are few choices. Recently, Zhang et al. (2018) constructed a component dictionary by combining six existing types of sentiment dictionaries to measure Weibo (microblogs in English, like Twitter in China) textual sentiment, which enlightened us in the text analysis task in this paper.[3] By using a novel dataset that contains massive articles about stock market analysis obtained from an investors' community, namely, Xueqiu (snowball in English, https://xueqiu.com/) forum, we construct four textual sentiment indices to measure investor sentiment based on dictionary-based methods. To overcome the shortage of dictionary-based methods, such as incomplete and updates slowly, we propose an integrated sentiment dictionary in the spirit of Zhang et al. (2018). In particular, we extend the Chinese emotional vocabulary ontology database by incorporating the Emoji dictionary and trending word dictionary. Importantly, we add the L&M dictionary to make the Chinese sentiment dictionary more financial.

A growing number of existing literature conduct investigations, showing that the investor sentiment extracted from social network platforms or finance websites can be used to forecast financial market movement (see, for example, Bollen et al., 2011; Behrendt & Schmidt, 2018; Zoen Git Hiew et al., 2019; Guégan & Renault, 2020). For instance, deriving the textual sentiment from Twitter or Weibo, Bollen et al. (2011) and Zoen Git Hiew et al. (2019) showed that the constructed textual sentiment index improves on forecasting stock closed price and returns, respectively. However, the social network platforms contain abundant irrelevant emotions from noninvestors. As a result, other text sources such as finance websites and investors' communities have been gained wide attention, such as Yahoo! Finance (Das & Chen, 2007) and StockTwits (Guégan & Renault, 2020). These studies provide some evidence that the predictive accuracy of stock price movement performs better after the textual sentiment is taken into account.

The textual sentiment extracted from social networks and news can explain stock volatility by linear analysis in the developed capital market (Antweiler & Frank, 2004; Johnman et al., 2018; Behrendt & Schmidt, 2018; Audrino et al., 2020). For example, Behrendt and Schmidt (2018) documented that the high-frequency

[1]One can obtain data at https://www.quandl.com/data/AAII/AAII_SENTIMENT-AAII-Investor-Sentiment-Data.

[2]The six proxies are the dividend premium, NYSE share turnover, the closed-end fund discount, the equity share in new issues, and the number and average first-day returns on IPOs.

[3]These dictionaries are the HowNet dictionary, network word dictionary, degree adverb dictionary, negative word dictionary, emoticon dictionary, and Chinese emotional vocabulary ontology database.

Twitter sentiment of individual-level stocks could affect intraday volatility. Exploiting text data obtained from Twitter and StockTwits, Audrino et al. (2020) constructed sentiment indicator by using the deep-machine learning spectrum awareness (MLSA) technique. They showed that sentiment variables can improve stock volatility prediction performance by using a heterogeneous autoregressive (HAR) model. Antweiler and Frank (2004) found that more messages (or bullish messages), posted on Yahoo! Finance and Raging Bull, on the day t implied greater market volatility on the day $t+1$ significantly. Johnman et al. (2018) provided evidence that business news sentiment was connected with the Financial Times Stock Exchange 100 (FTSE 100) stock index volatility, and they reported that negative (positive) sentiment increases (reduces) stock volatility. However, there is no consensus about whether textual sentiment can improve stock volatility forecasts. For example, a piece of adverse evidence was shown by Das and Chen (2007) who reported that the disagreement measurement could not explain volatility.[4] In addition, the others, such as Behrendt and Schmidt (2018) and Audrino et al. (2020), found that the predictive power of textual sentiment on stock volatility was significant but not economically significant. Our work differs from these very related studies in three respects. First of all, the feasibility of the predictive model is expanded. Existed evidence examines the predictive power of textual sentiment in developed capital markets based on the linear model such as vector autoregression (VAR) or heterogeneous autoregression (HAR). But it is not necessarily suitable to the emerging market for instance the Chinese mainland. Although both the linear and nonlinear relationship between the sentiment indices and future stock volatility are investigated, we find that using textual sentiment to forecast stock volatility in the Chinese market is more likely by a nonlinear style rather than a linear fashion. Secondly, in addition to the volatility model, we are also interested in the sources of textual sentiment. In particular, we concentrate on articles obtained from the investors' community rather than comments from the social media or investor platforms. Thirdly and importantly, the previous researches mainly focus on the model fitting and residual diagnosis. We further calculate the certainty equivalent return (CER) to show the economic value of textual sentiment in forecasting stock volatility. Indeed, we find that the predictability of one-half (two out of four) sentiment indices is meaningful from an economic point of view.

Volatility is an uncertainty measurement of the return of a financial asset. It is of great significance to measure it accurately for asset pricing, risk management, and hedging strategies (Brooks & Persand, 2003; Wang et al., 2006). For example, it is used as a required input value when calculating value at risk (Gospodinov et al., 2006). Traditionally, econometric models, such as generalized autoregressive conditional heteroskedasticity (GARCH) family (Bollerslev, 1986) and HAR family (Corsi, 2009), are the first choice to model volatility, whereas, they are used on very strict terms although they are concise and intuitionistic; on the other hand are the recurrent neural networks (RNNs), which are widely used in time series predicting problems because it can detect long-term persistence and nonlinear dependencies without any assumption on the distribution of the target variables (Bucci, 2020). As a type of RNNs, the long–short-term memory (LSTM) model is one of the most advanced ML architectures for sequence learning tasks (Fischer & Krauss, 2018). For example, Liu (2019) and Fischer and Krauss (2018) exploited LSTM model to forecast stock index volatility and stock movement, respectively. A very recent study of Bucci (2020) documented that the LSTM improved predictive accuracy of realized volatility comparing with existing econometric models. Many shreds of evidence have shown that the LSTM model is a state-of-the-art technique in financial time series forecasting. In this study, we exploit its nonlinear map ability to test the predictive power of sentiment indices on future stock volatility. Intuitively, we observe that the LSTM model can capture the dynamic feature of volatility well. Furthermore, based on data analysis and statistical tests, we show that the LSTM model can significantly examine the predictability of sentiment indices. The out-of-sample result illustrates that the sentiment models perform better than the benchmark model without sentiment. This finding suggests that the LSTM model is appropriate to excavate the forecastability of textual sentiment to stock volatility. In addition, we also consider the financial market heterogeneity that comes from the difference in time horizon. Thus, we discuss the changes of impacts of sentiment on stock volatility with different horizons. We find that the LSTM model and its sentiment models have an outstanding predictive performance with the horizon longer.

What distinguishes our work from others' on this subject include that we (i) focus on the linear and nonlinear relationship between stock volatility and textual sentiment in the emerging capital market, that is, the Chinese stock market; (ii) analyze textual sentiment by constructing a special Chinese sentiment dictionary based on a synthetic method and consider the memory persistence of investor when constructing sentiment indices; (iii) find

[4]The disagreement measure is defined as $\mathrm{DISAG} = |1-|(B-S)/(B+S)||$, where B and S are the numbers of buy and sell message on Yahoo! Finance, respectively.

that it is the nonlinear framework that is more appropriate to test the predictive impact of textual sentiment on stock volatility, which is a piece of new evidence that differs from the developed market; and (iv) investigate the economic value of textual sentiment.

The contributions of this study can be expressed as follows: Firstly, our result shows that the textual sentiment extracted from the Chinese investors' community affects stock volatility only by a nonlinear mechanism and that the textual sentiment can explain future stock volatility has been discussed by previous studies (see a few pieces of evidence, for example, Antweiler & Frank, 2004; Audrino et al., 2020; Johnman et al., 2018). Most evidence is from the developed capital market. Whether and how the textual sentiment from the emerging market affects stock volatility has not been effectively and adequately explored. Besides, most previous studies use a linear model to show the predictive power of sentiment. Discussion about nonlinear mechanisms remains an interesting direction. The linear forecast framework may be easy to understand, but it is not as flexible as the nonlinear model and maybe not always adaptive. Especially, comparing with the developed market, the Chinese stock market is more complex and chaotic owing to the immature structure of investors and regulatory policy. For example, the development of institutional investors is unbalanced, and many individual investors are irrational. The regulatory system is not perfect, and the degree of marketization is not high, and so on. In this study, both linear and nonlinear models are used to detect the predictive power of textual sentiment. We discover that the sentiment indices are the nonlinear cause of stock volatility. In particular, the stock volatility can Granger cause sentiment indices rather than vice versa. This is consistent with the findings of Wang et al. (2006) who use market data to measure investor sentiment. Besides, by using the nonlinear causality analysis framework that is based on the feed-forward neural networks, we find a significant causality from sentiment to volatility, As a result, the linear regression models are not suitable to test the predictability of sentiment indices to sock volatility in the Chinese market. To the best of our knowledge, this is different from the previous results.

Secondly, our work provides two new ideas in terms of the selection of text source and construction of sentiment index. First of all, we choose articles about stock market analysis rather than comments posted in an investors' community as textual sentiment sources, which enriches related studies in the selection of textual sentiment sources. Most of the studies exploit all comments published on social media or financial website platforms to derive the textual sentiment, for example, Zhang et al. (2018) and Behrendt and Schmidt (2018). However, they ignore the text noise because many comments contain irrelevant information like an advertisement but have an impact on human sentiment. To this end, the articles, written by experienced investors or analysts and read by a number of investors, as a source are exploited to extract textual sentiment because they are professional and with less noise. The second consideration is that we take investor memory into account when constructing the sentiment indices, which corresponds to reality more. Former researchers assume that investor sentiment is transient. Usually, they construct a statistical measurement based on intraday data to define textual sentiment directly. They think that the textual sentiment on day t has an impact on the stock market on day $t+1$ only, for example, Das and Chen (2007) and Johnman et al. (2018). However, we think that investor sentiment is persistent because investor memory is not negligible. For example, bad emotion yesterday may affect the emotion today. Yet, this influence is limited because individuals are also forgetful. In the empirical analysis, we use the minimum sentiment index to donate the worst sentiment during the past 3 days, and we find that this measure has a larger correlation with stock volatility. Moreover, we show that it outperforms in short-horizon volatility prediction and has larger economic values relative to the other sentiment measures.

Finally, our study explains that the textual sentiment can improve stock volatility prediction by using a nonlinear regression model, and this improvement is significant and makes economic sense. A large number of empirical studies show that the ML method performs far better than the econometric methods in volatility prediction (see, for example, Bucci, 2020; Liu, 2019). The nonlinear models especially the deep learning techniques are more suitable to the complex and chaotic financial system. In this study, we apply a state-of-the-art deep learning technique, the LSTM model, to investigate the nonlinear predictive power of investor sentiment measured by textual sentiment indices on stock volatility. We prove that the textual sentiment is a powerful forecasting factor. What is more, the performance of sentiment is investigated during the financial crisis. We demonstrate that sentiment is useful in terms of predictive power for short-horizon volatility rather than long horizon because sentiment is time-varying fast in this period. Furthermore, we show that predictability is still significant after controlling the trading volume variable, which is considered as information flow and sentiment proxy. In addition, we investigate the economic value of textual sentiment in forecasting stock volatility. We illustrate that one-half of textual sentiment indices, more specifically two-fourths, can improve the CER. These findings suggest that using textual sentiment to predict stock volatility is very meaningful and valuable. Importantly, this contributes to further understand what role does investor sentiment play in mean–variance equilibrium.

The remainder of the paper is organized as follows. Measurements and methodologies are presented in Section 2. Text data, text processing, and experiment setup are shown in Section 3. We show the empirical results including linear, nonlinear, and economic value analyses in Section 4. Finally, we make conclusions and discussions in Section 5.

2 | MEASUREMENT AND METHODOLOGY

This section introduces the principle of the LSTM model that is used to test the forecast power of the textual sentiment on stock volatility. Besides, we also show the integrated sentiment dictionary and sentiment measurements, which are used to extract textual sentiment in the next section.

2.1 | Volatility measurement

Volatility is a measurement of the uncertainty of financial asset return. There are several volatility definitions, such as realized volatility (Andersen & Bollerslev, 1998), absolute return (Ding et al. 1993), daily range (Alizadeh et al., 2002), the new statistical measurement using high, low, open, and closed price (Das & Chen, 2007),[5] and among others. The square root of the realized variance (RV) is used to define volatility in this study. Specifically speaking, letting p_t be closed price at trading day t, logarithmic return (log-return) of stock index on day t is defined as

$$r_t = 100 * (\log(p_t) - \log(p_{t-1})). \tag{1}$$

Then, volatility of the stock index is defined as

$$vol_t = \left[\frac{1}{\tau_t}\sum_{t-\tau_t}^{t}(r_t - \bar{r}_t)^2\right]^{1/2}, \tag{2}$$

where $\bar{r}_t$ represents the average on log-return during τ_t period before time t. Weekly, 2-week, and monthly volatilities refer to the situations when $\tau_t = 5$, 10, and 22, respectively. It is important to point out that forecasting volatility is different from predicting return. The former focuses on how stable the stock price will be in the future rather than how well the stock will be (the latter).

2.2 | LSTM model

The LTSM model is a special class of RNNs, which is made up of one input, one output, and one or more hidden layers. The most significant feature of this model is that the output of its hidden layer can be as an input to the next hidden layer, which means long memory. The LSTM model is first proposed in 1997 and interpreted intuitionally by Olah (2015). It becomes more popular in recent years due to the development of big data. The main structure and working principle are reflected by memory cells located in the hidden layer. Figure 1 shows how each memory cell works at timestep t. As illustrated in the picture, A is a chunk named memory cell of the neural network, which is similar to a workshop. Inside of A, S_t refers to a cell state, which is able to continuously update over time as information goes in or out. Each A contains three gates, ① the forget (f_t), ② input (i_t), and ③ output (o_t) gates , which are used to adjust cell state by filtering out information. In particular, the forget gate determines how much information should be removed. The input gate determines how much information should be added, and the output gate determines how much information should be used as output. x_t is independent variable (means new information), and v_{t-1} is output of last timestep. The workflow of each memory cell includes the following four steps.

Step 1: Forget gate determines how much information should be removed from the cell state at last timestep by sigmoid function. The information are from explanatory variable at t (i.e., x_t) and output of time $t-1$ (i.e., v_{t-1}). This process is presented by the equation,

$$f_t = \text{sigmoid}\left(W_{f,x}x_t + W_{f,v}v_{t-1} + b_f\right). \tag{3}$$

Obviously, f_t varies from 0 to 1, where 0 means discarding all the previous information and 1 represents keeping all the previous information.

Step 2: Partial information of x_t and v_{t-1} will be stored into the cell state. First, the input gate decides how much information will be added by using the formula:

$$i_t = \text{sigmoid}(W_{i,x}x_t + W_{i,v}v_{t-1} + b_i). \tag{4}$$

Next, the hyperbolic tangent function translates x_t and v_{t-1}

[5]They define volatility by using the following formula: $vol_t = 2\left(p_{t,high} - p_{t,low}\right)/\left(p_{t,open} + p_{t,closed}\right)$, where $p_{t,high}$, $p_{t,low}$, $p_{t,open}$ and $p_{t,closed}$ represent high, low, open as well as closed price, respectively.

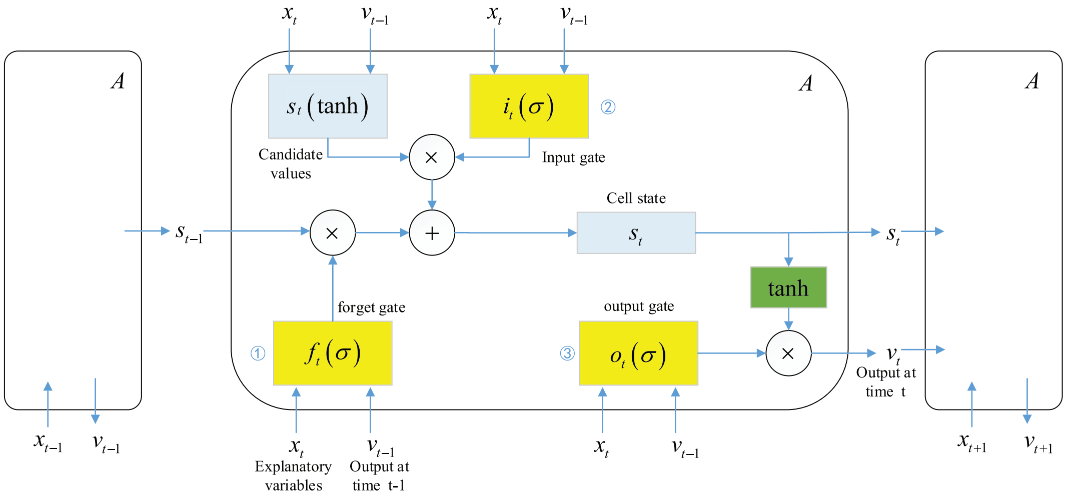


FIGURE 1 Structure of long–short-term memory (LSTM) networks memory cell following Olah (2015)

to be a candidate value of state through the following equation:

$$\tilde{s}_t = \tanh(W_{\tilde{s},x}x_t + W_{\tilde{s},v}v_{t-1} + b_{\tilde{s}}). \quad (5)$$

Step 3: Then, combining reserved information by using the forget gate and the input gate to update cell state though

$$s_t = f_t \circ s_{t-1} + i_t \circ \tilde{s}_t, \quad (6)$$

where ∘ represents Hadamard product.

Step 4: Finally, the output gate determines how much information of x_t and v_{t-1} will be sent to the next timestep by

$$o_t = \text{sigmoid}(W_{o,x}x_t + W_{o,v}v_{t-1} + b_o), \quad (7)$$

and then the output gate along with the updated cell state is used to produce final output v_t though

$$v_t = o_t \circ \tanh(s_t), \quad (8)$$

where W represents weight vector and b is bias vector.

2.3 | Dictionary-based sentiment analysis method

The ML and dictionary-based methods are two main textual sentiment analysis techniques. In this study, the dictionary-based method that is well understood, stable, and clearer is used to derive textual sentiment from textual articles. This method has several advantages, such as objective, easy to expand, and friendly sharing (Loughran & McDonald, 2016). Objectivity means that researchers can avoid being subjective once the sentiment word list is predetermined. Expansibility represents that one can expand word lists according to different research requirements. And sharing indicates that one could conduct further research work on the basis of previous studies. Importantly, the dictionary-based method can achieve similar accuracy as high as the ML method once one adopts an appropriate sentiment dictionary and tokenizer. For those reasons, we construct an integrated sentiment dictionary and several sentiment measurements for the textual sentiment analysis task.

2.3.1 | Integrated sentiment dictionary

As described in Figure 3 (below), the integrated sentiment dictionary that contains 12,218 negative words and 13,617 positive words consists of the following four lexicons:

(i) **Emoji dictionary** includes 59 Emoji on the comment bar of the Xueqiu forum. These specific emoticons are widely used by mobile clients. In order to obtain the polarity of Emoji, we invite five experienced investors who are users of Xueqiu forum to classify each Emoji into a category, that is, positive, neutral, and negative. According to the voting mechanism, a simple summary of Emoji polarity is shown in Figure 2.

(ii) **Trending word dictionary** is made up of 100 popular internet words during the 2013–2018 period

Emoji	Number	Polarity
	23	Positive
	10	Neutral
	26	Negative

FIGURE 2 Polarity of Emoji word list

concluded by Tencent. Trending words are easy to be accepted and used by the public. For each trend word, its sentiment polarity is determined manually based on its original meaning.

(iii) **DUT dictionary** refers to the Chinese emotional vocabulary ontology database.[6] It is a sentiment dictionary proposed by Dalian University of Technology (Xu et al., 2008) and widely used to Chinese text analysis (see, for example, Ren et al., 2018; Zhang et al., 2018).

(iv) **L&M dictionary** is a professional English sentiment dictionary in financial field proposed by Loughran and McDonald (2011). It is translated into the Chinese language by using machine translation techniques. We obtained L&M dictionary of Chinese language version including 421 positive words and 2,285 negative words.[7]

2.3.2 | Sentiment measurement

Antweiler and Frank (2004) define a simple statistical measure named "bullishness" based on the number of "relevant" messages and sum of messages of type c, where c could be call, hold, or sell messages. Similarly, letting N_p and N_n represent the number of positive and negative words on article i, respectively, we first define the sentiment index of article i as

$$S_i = \left(N_p - N_n\right) / \left(N_p + N_n\right). \tag{9}$$

Then the mean sentiment index on day t, Mean$_S$enti (S_t), is represented by the average of S_i within day t. That is,

$$\text{Mean\_Senti}_t = S_t = \sum_{i=1}^{n_t} S_i / n_t, \tag{10}$$

where n_t refers to the total number of published articles within day t. We can see that, on the day t, sentiment tends to be more positive if the mean sentiment index is closed to 1 whereas more negative if it approaches -1.

However, the mean sentiment index is constructed based on the number of sentiment words, which ignores human memory. As we all know, the memory of an investor is persistent but not too long because humans are also forgetful. For example, one is not happy about something yesterday; maybe he or she is able to remember until tomorrow. On the basis of this idea, we propose the following three sentiment indices by exploiting the mean sentiment index of three consecutive days, which are the maximum sentiment index,

$$\text{Max\_Senti}_t = \max(S_t, S_{t-1}, S_{t-2}), \tag{11}$$

the minimum sentiment index,

$$\text{Min\_Senti}_t = \min(S_t, S_{t-1}, S_{t-2}), \tag{12}$$

and the average sentiment index,

$$\text{Avg\_Senti}_t = (S_t + S_{t-1} + S_{t-2})/3. \tag{13}$$

In the empirical analysis, we use these sentiment indices on the day t to predict stock volatility on the next trading day, that is, $t+1$. Under this framework, we remove the sentiment index on some days in order to match trading days.

3 | TEXT DATA, TEXT PROCESSING, AND EXPERIMENT SETUP

We introduce an active investors' community in the Chinese market, namely, Xueqiu forum, which is the resource of the textual sentiment. Text processing and experiment setup in the empirical analysis are also presented in the section.

[6]See https://ir.dlut.edu.cn/EmotionOntologyDownload.

[7]We use the latest version of 2018; see https://sraf.nd.edu/textual-analysis/resources/#LM%20Sentiment%20Word%20Lists. Note that the original L&M dictionary includes 354 positive words and 2355 negative words. It differs from the Chinese language version because an English word may correspond to different Chinese words that are synonyms. As for the machine translation technique, we exploit the top three most popular software, which are Google, Baidu, and Youdao Translators. We also check the translation result manually.

3.1 | Text data

We choose the Xueqiu investors' community as a resource of text data. Xueqiu is established in March 2010. It is a good source of textual sentiment for three reasons. Firstly, Xueqiu has a large number of users. Up to July 2015, registered users of Xueqiu have already exceeded 6 million, and the number increases to more than 10 million in the mid of 2018. Secondly, most of the users are highly educated and relatively rich. Around the composition of users as of 2015, 71.94% of them have bachelor's degrees or above (up to 76.20% in 2018); 64.54% come from first-tier cities of China, such as Beijing and Shanghai, and 65.60% are between 30 and 45 years old, indicating that the majority of them have sufficient disposal funds to invest. And thirdly, Xueqiu is an active investors' community. According to the platform's mid-year data report 2018, the average daily online time of users is 48 min. In addition, this platform produces an average of 200,000 discussion records on financial hotspots per day, and the top three popular articles are read by as high as 16.314 million, 10.537 million, and 9.602 million users, respectively. However, to the best of our knowledge, there are rare researchers who focus on Xueqiu except for Zhang et al. (2018).[8]

3.2 | Text processing

Text processing is the center of textual sentiment analysis. As shown in Figure 3 (below), three steps are included as follows: text data obtaining (web crawling), text data clearing, and text data transforming. In the process of data transforming, we exploit the integrated sentiment dictionary and the constructed sentiment measurements to transform the text data into a numerical format.

3.2.1 | Web crawling

We download the main article information of the Xueqiu investors' community through web crawling. The obtained text information includes the author's names, the number of followers of the author, posted time, title, contents, weblink, and read count of the published article. Note that it is difficult to obtain the article contents directly because the unique identifier of each article is ruleless. To address it, we crawl all links of the articles first and then obtain articles' information such as titles and contents through these links.

3.2.2 | Data cleaning

We aim to analyze more useful and valuable text information, and therefore, noise reduction is what we concentrate on. Firstly, some articles that are under review and not visible are removed. Secondly, duplicate articles are also deleted according to the unique article link. In addition, we discard the published articles written by users with less than 30 followers. This is because fewer followers mean lesser attention and lesser impact. Finally, we obtain about 80,000 articles for this study.

3.2.3 | Data transforming

We exploit the integrated sentiment dictionary and constructed sentiment measurements described in Sections 2.3.1 and 2.3.2 to transform document-level text data into numerical format. This process mainly consists of two parts: word segmentation and statistics modeling, including the following four steps. (i) Unlike the English word naturally separated by blank, we need to choose a segmentation tool to segment the sentences into separate words. A Chinese language text segmentation technique, *jieba*, is used for this task because it can improve accuracy when privative appears.[9] In addition, the exact pattern of the *jieba* method is chosen in that it is accurate and suitable for text analysis. (ii) The split words are projected onto the sentiment dictionary space to calculate the number of positive and negative words. (iii) The sentiment scores of each document are obtained based on the sentiment measurements. And (iv) we obtain daily sentiment time series by gathering the sentiment score of each article on each day.

3.3 | Experiment setup

The textual sentiment data and stock volatility data are at daily frequency, 1230 observations in total; 67% of the data set (830 days) is used for training and the left (400 days) for testing. When training the LSTM model,

[8]Our work is fundamentally different from their work mainly including three aspects. Firstly, they use textual sentiment to forecast stock price up or down, but we predict the volatility value. Secondly, their methods consist of a multilayer perceptron and support vector machine, whereas we use the LSTM technique. Finally, they use stock comment text posted by all kinds of the participant over half a year period, yet we make use of articles about stock market analysis around 5-year period posted by the influential person (has more than 30 fans).

[9]For the *jieba*package, see https://github.com/fxsjy/jieba. It is described in both English and Chinese languages.

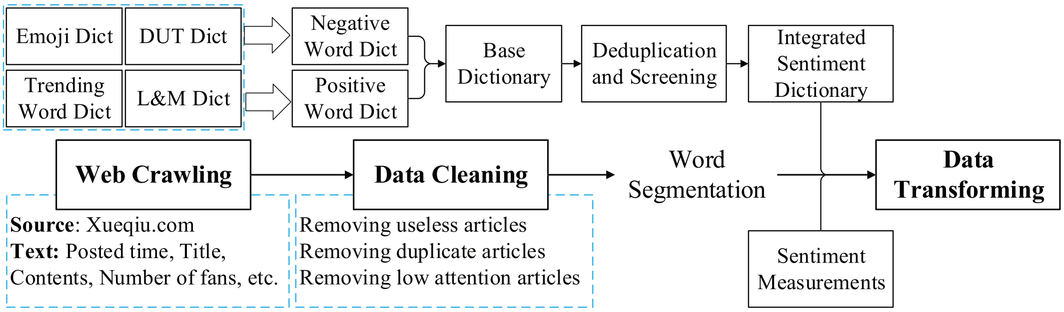


FIGURE 3 Construction of the integrated sentiment dictionary and text processing

we employ rolling windows and set its look back to five. Moreover, we standardize input data by MinMaxScaler. When standardizing, we process training and test sets separately in order to avoid information leakage.

Our LSTM network includes one LSTM layer with fifteen neurons and a full connection layer with one neuron. In addition, we use a linear activation function for the fully connected layer. The compilation optimizer is Adma optimizer (Kingma & Ba, 2014). We set the learning rate as 0.0001 and choose mean squared error (MSE) as the loss function over the training period. Web crawling, text data parsing, clearing, storing, and transforming are conducted on Python 3.7, based on the packages pandas, etree, BeautifulSoup, pymongo, *jieba*, and numpy. The LSTM model is constructed based on Keras with TensorFlow as backend and trained on NVIDIA GPU.[10] Other data analysis relies on EViews 10 and R 3.6.3.

4 | EMPIRICAL ANALYSIS

The empirical results are presented in this section. Firstly, we introduce the sources and periods of data and analyze the data feature based on descriptive statistics. Secondly, we investigate the impacts of sentiment indices on future stock volatility by using both linear and nonlinear frameworks. In particular, we explore the correlation, Granger causality, and nonlinear causality relationships between sentiment indices and stock volatility. We also examine the predictive power of sentiment indices on future stock volatility with different horizons by exploiting the LSTM model during the full sample. Besides, it was well-known that the Chinese stock market turbulence that a third of the value of A-shares was lost within 1 month happened in 2015. According to Wikipedia, the period from June 12, 2015, to February 1, 2016, can be viewed as a financial crisis stage in the Chinese stock market, and we also investigate the sentiment performance during this period. Furthermore, we study the predictability of sentiment indices after controlling the trading volume variable that is identified as a proxy of sentiment. Finally, we show the economic meaning that using textual sentiment to predict stock volatility.

4.1 | Data sources

The articles, written by experienced investors or analysts from the Xueqiu investors' community, are the source of textual sentiment. They are located in the Shanghai–Shenzhen stock market plate of the Xueqiu platform, which can represent market sentiment well.[11] We determine the study period starting from March 31, 2014, because the Xueqiu platform comes into Vogue from this time. As a capitalization-weighted stock market index, the Shanghai–Shenzhen 300 stock index (CSI 300 for short) covers about 60% of the market capitalization of A-shares, which could be a good proxy of the Chinese stock market. The stock price data come from the Wind database. According to the period selection of text data, the time spans from March 31, 2014, to April 12, 2019.

4.2 | Summary description

The sequence diagram of log-return is shown in Figure 4. The log-return is stationary according to the augmented

[10]More details about software and hardware can be concluded as follows: Windows 10 64-bit, CUDA v-10.0, cuDNN v-7.6.5, python v-3.7.5-amd64, Keras v-2.2.2, TensorFlow-gpu v-1.13.1, and GPU hardware: GeForce GTX 1600 Ti.

[11]The Shanghai–Shenzhen, Hong Kong, and American stock markets are three main plates in Xueqiu forum. However, we find that the plate setup is canceled in 2020 and replaced by a whole plate. We obtain the text data in April 2019.

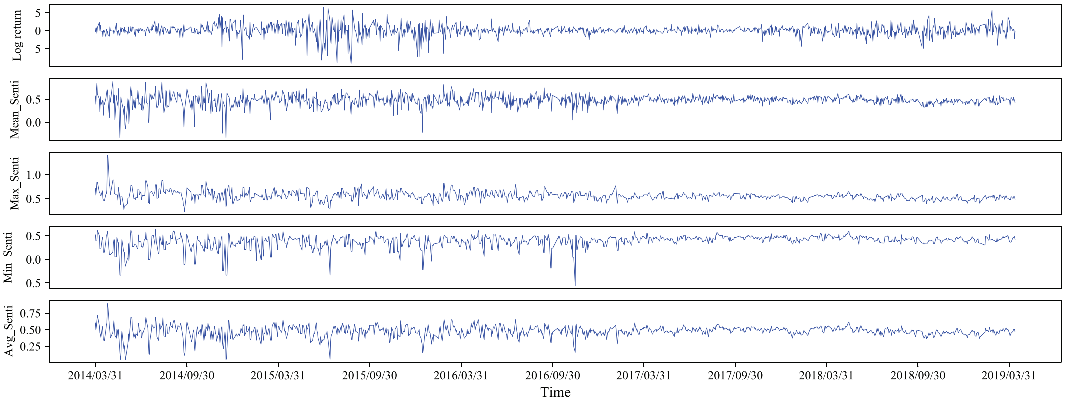


FIGURE 4 Time series of return and sentiment indices

Dickey-Fuller (ADF) test with a t statistic of -33.419. The statistical analysis of log-return, volatility, and sentiment indices is described in Table 2. We find that the standard deviation of volatility goes smaller while the mean value of volatility goes bigger with the longer horizon. Besides, the skewness values indicate that log-return is negative skewness while volatility follows a heavily right-skewed distribution. In addition, the reported kurtosis shows that both log-return and volatility exhibit positive excess kurtosis. Moreover, the Q statistics of the Ljung–Box test suggests that both log-return and volatility are dynamic dependencies. What is more, most of the variables are stationary according to the ADF test. Whereas, the monthly volatility is not stationary, which implies that the econometric model is improper to make statistical inferences. Finally, we observe that neither log-return nor volatility follows a normal distribution because the Shapiro–Wilk test (SW test) rejects the null hypothesis.

The statistical summary and time series graph of four sentiment indices defined by Equations (10)–(13) are presented in Table 2 and Figure 4, respectively. From the table, we observe that all mean values are positive. It is not surprising because the textual sentiment is extracted from rational articles about stock market analysis. Besides, sentiment indices exhibit negative skewness except for the maximum sentiment index. We also find that all of them show positive excess kurtosis and dynamic dependencies. The reported ADF statistics indicate that all the sentiment time series are stationary.

4.3 | Linear analysis

This subsection investigates the predictive power of sentiment indices on stock volatility by using a linear analysis framework. In particular, we discuss the correlation and causality relationship based on the Pearson correlation analysis and the Granger causality test, respectively, which aims to primarily explore whether sentiment indices have a predictive impact on stock volatility from a linear point of view.

4.3.1 | Correlation analysis

Panel A of Table 3 presents the contemporaneous correlation between stock volatility and sentiment indices during the full sample. First of all, it shows that sentiment indices are significantly negatively related to stock volatility,[12] which suggests that stock volatility is more sensitive to negative sentiment. This is in line with the news that we record at the beginning of the paper. Facing the same event, that is, China–U.S. trade war, the A-shares market has different reactions to bad news and good news (6% up for negative affair vs. 2% down for positive affair). Secondly, with the volatility horizon being longer, the correlation between the mean sentiment index and volatility goes weaker for the full period. Whereas, that relation goes strong first and then turns weak for other sentiment indices. Finally, we note that the absolute value of the significant correlation coefficient varies from 0.075 to 0.186. A similar result is presented by Wang et al. (2006) who use five proxy indicators, the put−call volume ratio (PCV), put−call open interest ratio (PCO), Short-Term Trading Index (or say, ARMS ratio),[13] AAII, and II, to measure investor sentiment. They show that the absolute value of the correlation coefficients between volatility and sentiment index changes from 0.063 to 0.175 for the full sample

[12]We also investigate the correlation between stock volatility and the changes in sentiment indices, but we obtain insignificant results.
[13]The ARMS ratio is defined as the number of advancing issues scaled by the trading volume (shares) of advancing issues divided by the number of declining issues scaled by the trading volume (shares) of declining issues.

TABLE 2 Descriptive statistics of return, volatility, and sentiment indices

Variables	Mean	Med	Max	Min	*SD*	Skew	Kurt	*Q*(5)	*Q*(20)	ADF	Shapiro–Wilk
Log-return	0.050	0.069	6.499	−9.154	1.564	−0.969	9.249	1.370e1**	1.130e2*	−33.419*	0.896 (0.000)
Weekly volatility	1.234	0.992	6.690	0.130	0.943	2.129	8.589	3.395e3*	8.435e3*	−4.041*	0.795 (0.000)
2-week volatility	1.294	1.064	5.295	0.232	0.863	1.794	6.732	5.126e3*	1.406e4*	−2.863***	0.832 (0.000)
Monthly volatility	1.326	1.065	4.491	0.265	0.812	1.550	5.375	5.889e3*	1.994e4*	−1.889	0.852 (0.000)
Mean_Senti	0.478	0.486	0.889	−0.333	0.122	−1.399	10.104	7.138e1*	1.230e2*	−28.997*	0.907 (0.000)
Max_Senti	0.561	0.553	1.400	0.231	0.093	1.593	14.097	7.606e2*	1.036e3*	−18.593*	0.915 (0.000)
Min_Senti	0.386	0.419	0.634	−0.556	0.137	−2.336	11.146	6.450e2*	1.118e3*	−10.504*	0.805 (0.000)
Avg_Senti	0.477	0.486	0.895	0.048	0.084	−1.088	7.965	6.397e2*	7.633e2*	−17.733*	0.919 (0.000)

Note: This table shows the descriptive statistics of stock log-return, volatility, and sentiment indices. $Q(n)$ refers to Q statistics of the Ljung–Box test with n lags, which test the null hypothesis that time series is not autocorrelation. The ADF represents the statistics of the augmented Dickey–Fuller test of which null hypothesis is that a unit root is present in a time series sample. The Shapiro–Wilk test is reported in the last column, and the p values that are less than 0.01 are shown in the round bracket, which implies that all series do not follow the normal distribution. [Correction added on 11 May 2021, after first online publication: In Table 2, the citations of statistical significance at 1% level have been corrected.]

*Represents statistical significance at 1% level.

**Represents statistical significance at 5% level.

**Represents statistical significance at 10% level.

TABLE 3 Correlation, linear, and nonlinear Granger causality analysis

Correlation/null hypothesis	Period	Mean_Senti	Max_Senti	Min_Senti	Avg_Senti
Panel A: contemporaneous correlation between volatility and sentiment					
Correlation between weekly volatility and sentiment	Full sample	−0.140	−0.075	−0.156	−0.163
	Financial crisis	−0.319	−0.403	−0.254	−0.402
Correlation between 2-week volatility and sentiment	Full sample	−0.128	−0.087	−0.179	−0.186
	Financial crisis	−0.271	−0.364	−0.226	−0.374
Correlation between monthly volatility and sentiment	Full sample	−0.086	−0.046	−0.140	−0.131
	Financial crisis	−0.110	−0.164	−0.131	−0.182
Panel B: linear Granger causality analysis					
Sentiment does not cause weekly volatility linearly	Full sample	3.829***	0.202	0.000	0.344
	Financial crisis	0.009	4.067**	0.264	1.101
Sentiment does not cause 2-week volatility linearly	Full sample	0.255	0.411	0.305	0.010
	Financial crisis	1.125	1.608	0.001	0.575
Sentiment does not cause monthly volatility linearly	Full sample	0.084	0.074	0.015	0.011
	Financial crisis	2.711	1.767	1.716	2.475
Panel C: nonlinear causality analysis					
Sentiment does not cause weekly volatility nonlinearly	Full sample	684.421*	351.456*	614.656*	585.645*
	Financial crisis	24.173*	25.264*	23.394*	28.524*
Sentiment does not cause 2-week volatility nonlinearly	Full sample	1028.370*	581.943*	802.124*	919.785*
	Financial crisis	9.753*	9.208*	9.342*	13.379*
Sentiment does not cause monthly volatility nonlinearly	Full sample	918.055*	531.324*	711.373*	851.147*
	Financial crisis	4.158*	2.161*	7.083*	11.165*

Note: Panel A of this table reports the contemporaneous correlation between stock volatility and sentiment indices. Panels B and C report the *F* statistics of linear and nonlinear Granger causality test, respectively. In particular, the linear causality analysis is based on the work of Granger (1980). The nonlinear causality is an extension of the linear causality based on the feed-forward neural networks.
*Represents statistical significance at 1% level.
**Represents statistical significance at 5% level.
**Represents statistical significance at 10% level.

(daily level data) and from 0.008 to 0.270 for the full sample (weekly level data). Besides, Huang and Liu (2020) also document that the sentiment scores have a negative correlation of 0.186 with the stock price change.

The correlations between volatility and sentiment during the financial crisis are also shown in Panel A of Table 3. Generally, we find that the correlation gets stronger in this period. The absolute value of the coefficient varies from 0.110 to 0.403, which is significantly larger than that during the full sample. This suggests that investors are more emotional during the depression period, and thus, sentiment has a larger correlation with stock volatility. In addition, we observe that correlation goes weaker with the volatility horizon being longer for all sentiment indices.

4.3.2 | Granger causality analysis

Correlation is fundamentally different from causal relation. Based on the Granger (1980) causality theory, a linear causality relationship between stock volatility and sentiment indices is explored. The results as reported in Panel B of Table 3, broadly speaking, suggest that the sentiment indices cannot Granger cause volatility whether it is during the full sample or financial crisis. On the contrary, we find significant evidence that volatility can Granger cause sentiment in most cases. We discuss different lag timesteps from 1 to 5 and get consistent conclusions. These results are not reported because of space limitation but are available on request. Our results confirm the findings of Wang et al. (2006) who find that the

AAII, II, PCV, and PCO sentiment indices are affected by stock index volatility, rather than vice versa. As a result, linear causality analysis shows that stock volatility is a cause of sentiment indices, rather than an effect.

4.4 | Nonlinear predictive analysis

No evidence shows that the sentiment indices can affect stock volatility by using a linear regression framework. The nonlinear pattern is investigated in this subsection. Firstly, a nonlinear causality test is conducted. Then, we apply a state-of-the-art deep learning method, namely, LSTM model, to investigate the predictive power of sentiment indices on stock volatility. To examine the effectiveness of the sentiment indices, we build a benchmark model that considers the volatility history information only in the LSTM model. The sentiment model refers to that the sentiment index as an additional independent variable is considered into the benchmark model. To evaluate the significance of the predictability of sentiment indices, we consider two loss functions and the out-of-sample R^2 to compare the prediction accuracy between the benchmark model and the sentiment models.

4.4.1 | Nonlinear causality analysis

To investigate whether there exists nonlinear predictability in the sentiment indices, we first detect the nonlinear causality relationship between the stock volatility and sentiment indices. As an extension of the Granger causality test, the nonlinear causality test is based on the feed-forward neural networks. Recently, Hmamouche contributes an R package named "NlinTS" to realize it.[14] The results are shown in Panel C of Table 3. Fortunately, we see that the F statistics are significant at 1% level, which means that the sentiment indices can cause stock volatility by a nonlinear framework. It is consistent with the result during the financial crisis even though it has obvious smaller Fstatistics. On the other side, the nonlinear causality from volatility to sentiment is also discussed. We find that volatility cannot cause sentiment through a nonlinear mechanism except for the average sentiment index. We omit the details owing to space limitations. These findings illustrate that it is reasonable to consider a nonlinear mode to investigate the predictive power of sentiment.

[14]See https://cloud.r-project.org/web/packages/NlinTS/index.html. In the empirical analysis, the parameter setting is concluded as follows: lag = 1, iters = 20, learningRate = 0.1, algo = "amda."

4.4.2 | Evaluation function

We use two types of loss functions, namely, the root mean squared error (RMSE) and the mean absolute error (MAE), to evaluate out-of-sample performance. They are basic and common criteria for comparing the predictive performance among ML methods (see, for example, Henrique et al., 2019). These measures are defined by

$$\mathrm{RMSE} = \left[\left(\frac{1}{T}\right)\sum\left(\tilde{v}_{pt} - v_{rt}\right)^2\right]^{1/2}, \tag{14}$$

and

$$\mathrm{MAE} = \left(\frac{1}{T}\right)\sum\left|\tilde{v}_{pt} - v_{rt}\right|, \tag{15}$$

respectively, where $\tilde{v}_{pt}$ and v_{rt} refer to the fitted value and the real value, respectively. T denotes the size of observations during out-of-sample. Note that the smaller loss function does not always mean the better prediction performance of the model will be. As a result, the Diebold and Mariano (2002) test statistic (D–M test) is used to compare predictive accuracy. The null hypothesis is that the expected loss of forecast error is equal. Thus, this test is based on the loss function. In order to compare the predictive accuracy between benchmark model and sentiment models, we conduct the D–M test to the two loss functions mentioned above.

In addition, we also use popular criterions in economics and finance field, that is, the Campbell and Thompson (2008) out-of-sample R^2, R^2_{OS}, to evaluate the prediction performance. The specification of R^2_{OS} is expressed as

$$R^2_{OS} = 1 - \sum_{t=1}^{T}\left(v_{rt} - \widehat{SV_t}\right)^2 \sum_{t=1}^{T}\left(v_{rt} - \tilde{}BV_t\right)^{-2}, \tag{16}$$

where $\widehat{SV_t}$ is the fitted value from the sentiment models estimated through period $t-1$ and $\widetilde{BV}_t$ refers to the fitted value from benchmark model estimated through period $t-1$. Intuitively, the positive R^2_{OS} shows that the sentiment model has lower mean squared prediction error than the benchmark model.

4.4.3 | Predictability of sentiment index

Table 4 presents the main findings of out-of-sample forecasting performance. By comparing different sentiment models with the benchmark model, we can conclude as follows. Firstly, for the benchmark model, we find that

TABLE 4 Out-of-sample predictive accuracy without trading volume

Horizon	$\tau_t = 5$			$\tau_t = 10$			$\tau_t = 22$		
Error accuracy	RMSE	MAE	R^2_{OS} (%)	RMSE	MAE	R^2_{OS} (%)	RMSE	MAE	R^2_{OS} (%)
Benchmark model	0.353	0.248	–	0.202	0.146	–	0.126	0.086	–
Mean_Senti model	0.347	0.241	3.26	0.194	0.138	7.19	0.122	0.082	6.66
	(4.201*)	(5.893*)		(5.599*)	(7.567*)		(6.379*)	(8.109*)	
Max_Senti model	0.349	0.245	2.13	0.200	0.144	2.30	0.125	0.085	2.09
	(5.117*)	(4.964*)		(4.867*)	(5.218*)		(6.556*)	(8.656*)	
Min_Senti model	0.344	0.237	4.90	0.195	0.138	6.90	0.123	0.082	6.11
	(5.051*)	(7.648*)		(5.175*)	(7.273*)		(6.650*)	(7.201*)	
Avg_Senti model	0.346	0.240	3.80	0.196	0.140	5.31	0.123	0.083	4.77
	(5.267*)	(7.228*)		(5.003*)	(6.331*)		(6.600*)	(8.242*)	

Note: This table reports the results of out-of-sample prediction, including the RMSE, MAE, and out-of-sample R^2. The Diebold and Mariano (2002) test (D–M test) statistics is shown in the round brackets. The alternative hypothesis of D–M test is that predictive accuracy of sentiment model is more accurate than that of the benchmark model. The benchmark model represents that the LSTM model includes volatility history information only, whereas the sentiment model denotes that the corresponding sentiment index is added into the benchmark model.
Abbreviations: MAE, mean absolute error; RMSE, root mean squared error.
*Represents statistical significance at 1% level.

the out-of-sample prediction performs better with volatility horizons being longer. The RMSE reduces from 35.36% to 20.18% and then to 12.64%. It is natural because longer horizon volatility means lower standard deviation (see Table 2), and the nonlinear model LSTM is easier to capture flat data feature. This finding is in line with the results of sentiment models. Note that this result is very meaningful. A famous financial market theory so-called heterogeneous market hypothesis states that the heterogeneity exists across traders. The sources of heterogeneity are manifold, such as institutional constraints and temporal horizons. Taking the heterogeneity into consideration helps to understand financial market behavior. For example, Corsi (2009) proposes a HAR-RV model based on the difference of time horizon and is widely cited by scholars. Our study shows that the LSTM model has a better-fitted performance in a longer horizon, which suggests it is a better choice for long-term investors.

Secondly, from the results of the D–M test and out-of-sample R^2, we identify that the sentiment indices improve significantly stock volatility forecasts based on the nonlinear model. This result, to some extent, confirms the work of Zoen Git Hiew et al. (2019) who report that the textual sentiment extracted from Weibo predicts the individual stock return time series more likely in a nonlinear fashion. On the whole, among these sentiment models, the Mean_Senti and Min_Senti models perform better on out-of-sample prediction relative to other models. Meanwhile, the optimal predictive model changes with the volatility horizon changing. For example, for weekly volatility (i.e., $\tau_t = 5$), the Min_Senti model has smallest loss function and largest R^2_{OS}. Comparing with the benchmark model, it improves predictive accuracy by 2.50% of RMSE, 4.64% of MAE, and 4.90% of R^2_{OS}. Interestingly, the Min_Senti model has a smaller value of loss function and greater R^2_{OS} than the Mean_Senti model, which examines that the consideration of investor memory is meaningful. This result indicates that the relatively down sentiment will continue to influence stock volatility in the short term. For 2-week and monthly volatilities (i.e., $\tau_t = 10$ and $\tau_t = 22$, respectively), the prediction accuracy is very closed for both of the Mean_Senti and Min_Senti models in terms of loss functions. Whereas, the Mean_Senti model outperforms slightly the Min_Senti model in terms of R^2_{OS}. Overall, the Min_Senti model is better for short horizon, whereas the Mean_Senti model is greater for medium and long horizons. These results suggest that investor memory is not negligible for short-term stock volatility prediction.

Thirdly, we observe that the Max_Senti model improves prediction accuracy a little relative to the benchmark model for all volatility horizons. This might be on account of weak correlation relationship as described in Table 3. When the volatility horizon changes from weekly to monthly, the absolute correlation coefficients between volatility and the maximum sentiment index are very low with ranging from 0.046 to 0.087. Whereas, the Max_Senti model still performs better with respect to the benchmark mode. A possible interpretation is that partial prediction information in the maximum sentiment index is identified and exploited by the LSTM model. Finally, we observe that the average sentiment index is also effective for stock volatility prediction.

In order to observe the prediction difference intuitively before and after adding the sentiment indices, Figure 5 displays the actual data (the blue one), train data (the green one), and prediction data (the red one) of different horizon volatility. On the left column, the benchmark models for weekly, 2-week, and monthly volatility forecasting are presented. We see that the LSTM model can draw the general trend of volatility only based on historical data. On the right column, the best predictive models in terms of the smallest RMSE value are shown. We find that the impacts of sentiment indices are intuitively reflected in the "turning point" of the forecasting diagram. That is to say, the sentiment index is able to capture the changes when volatility rises or declines suddenly due to the influence of events such as serious negative or positive news.

4.4.4 | Predictability during financial crisis

How does the textual sentiment perform out-of-sample during the financial crisis might be more attractive to readers.[15] To this end, we shorten our research period to investigate the out-of-sample predictive ability of sentiment in the depression period. In particular, we train the model by using in-sample data from March 31, 2014, to June 11, 2015, and thus, the out-of-sample dataset covers the financial crisis (from June 12, 2015, to February 1, 2016). The out-of-sample predictive accuracy is shown on Table 5. Firstly, we observe significantly larger values of RMSE and MAE in this period relative to the full sample. It is reasonable in that we use in-sample data with small fluctuations to train the model while use data with larger fluctuations to conduct out-of-sample fitting. As a result, the trained model is difficult to capture the dynamic of volatility in the financial crisis period. Another, as we all know, ML models are highly dependent on the amount of data. One may think that this result is caused by overfitting because our in-sample size is relatively small (285 in total). To address this issue, we shrink our model scale by controlling the number of neurons and investigate the out-of-sample performance. Unfortunately, similar results are obtained.

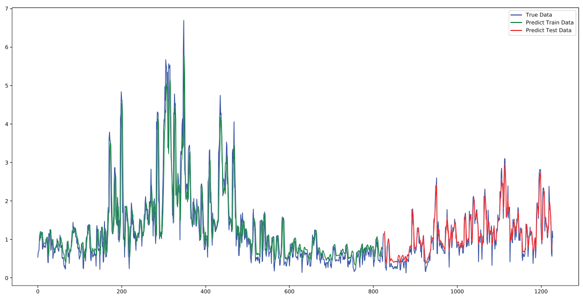
In addition, our empirical results show that the predictability of sentiment relies on the volatility horizon. In particular, for weekly volatility, we find that the sentiment models have significant improvement in terms of predictability with respect to the benchmark model. However, the out-of-sample performance of sentiment models gets poor as the horizon increases. For example, for 2-week volatility, the sentiment models have very closed values of the loss function with the benchmark model. While for the monthly volatility, the benchmark model even outperforms the sentiment model significantly. We account for this case from the following aspect. Volatility with longer horizon means that it has relatively smoother dynamic (see Figure 5). While during the financial crisis, the sentiment is time-varying with a large fluctuation, which can be proved by the larger standard deviation relative to the full sample based on the unpublished results. Hence, investors' behaviors are more complex and unpredictable in this period, and thus, it is easy to be reflected in a short time rather than a long term. In this way, sentiment indices have significant predictability for short-term volatility, that is, weekly volatility. On the contrary, long-term volatility contains longer history information and so that it is difficult to predict by short-term changeable sentiment because sentiment is time-varying fleetly especially during the financial crisis.

4.4.5 | Predictability after considering another sentiment variable

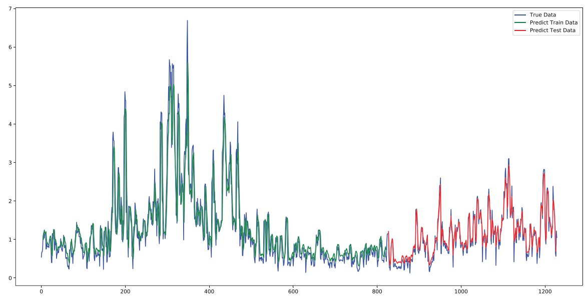
Trading volume is regarded as a proxy for the information flow. Some researchers also document that it is a good proxy of investor sentiment (Deeney et al., 2015; Uygur & Taş, 2014). Enlightened by their works, we also investigate the predictive power of sentiment indices after controlling trading volume. As displayed in Table 6, the benchmark model refers to that the trading volume and volatility history information as controlling variables are considered into the LSTM model, which differs from the previous benchmark model. And the sentiment model indicates that the corresponding sentiment index is added into this benchmark model. Several conclusions can be summarized as follows.

First of all, the results show that the trading volume has significant predictive power on stock volatility. For example, the values of RMSE and MAE decrease for all models relative to the situations without trading volume (comparing benchmark model from Tables 4 and 6). In addition, we find that the sentiment indices show a relatively outstanding predictive performance after adding the trading volume variable. In particular, the sentiment models have significantly lower RMSE and MAE relative to the benchmark model. Furthermore, the R^2_{OS} lies between 5.42% and 6.91% for weekly volatility (between 22.27% and 25.32% for 2-week volatility and between 13.75% and 15.38% for monthly volatility), which increases a lot comparing with the results shown in Table 4. These findings suggest that the trading volume enhances the predictive power of

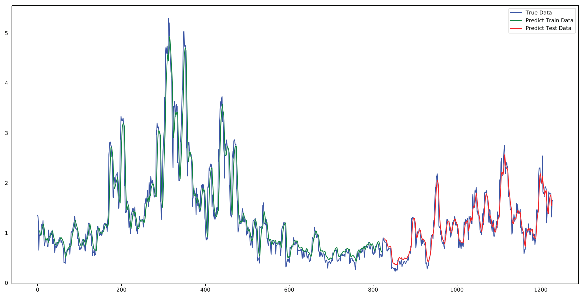
[15]We thank the anonymous reviewer for this valuable opinion.



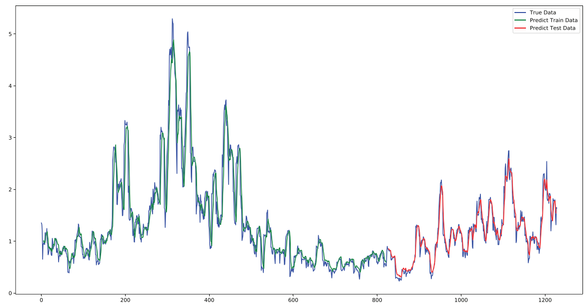
(A) Benchmark model when τ_t=5.



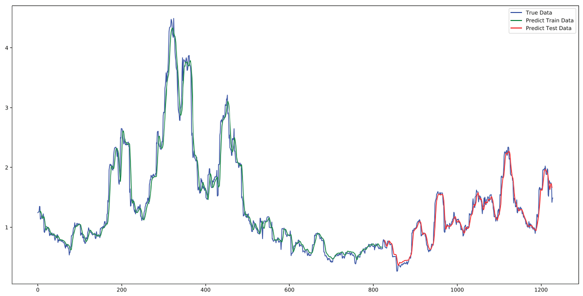
(A*) Min_Senti model when τ_t=5.



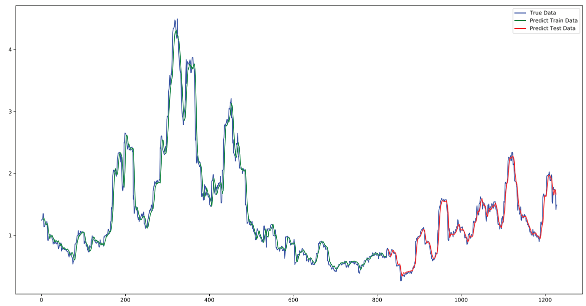
(B) Benchmark model when τ_t=10.



(B*) Mean_Senti model when τ_t=10.



(C) Benchmark model when τ_t=22.



(C*) Mean_Senti model when τ_t=22.

FIGURE 5 Prediction results of benchmark model and improved sentiment model with the smallest root mean squared error (RMSE). Actual, train, and prediction data are depicted by the blue, green, and red lines, respectively. (a) represents benchmark model when input variable only includes history information of weekly volatility, and (a*) adds Min_Senti index into model (a). Similarly, (b) denotes benchmark model of 2-week volatility, and (b*) adds Mean_Senti index into model (b). (c) is benchmark model of monthly volatility, and (c*) adds Mean_Senti index into model (c)

sentiment indices. By comparing different sentiment models, we find that the Avg_Senti model performs better than the others in weekly volatility prediction while the Mean_Senti model shows a stronger predictive performance in 2-week and monthly volatilities in terms of R^2_{OS}. Whereas, we see that the loss function values are very closed to each other for sentiment models. In summary, we conclude that the sentiment indices improve stock volatility prediction based on a nonlinear framework. And this predictive power is still strong after controlling the trading volume variable, which represents an information flow and sentiment proxy in the stock market.

4.5 | Economic values of sentiment index

The investigations above are based on data analysis. In this subsection, we will discuss the influence of sentiment indices from the economic value perspective. Following Campbell and Thompson (2008), Wang et al. (2016), and among others, we consider a mean–variance investor and calculate the realized utility gains to compare the economic values of different volatility models. More specifically, an investor allocates his/her asset to stock and risk-free treasury bills. In order to obtain more utility, an optimal weight in a risky asset is

TABLE 5 Out-of-sample performance during the financial crisis

Horizon	$\tau_t = 5$			$\tau_t = 10$			$\tau_t = 22$		
Error accuracy	RMSE	MAE	R^2_{OS} (%)	RMSE	MAE	R^2_{OS} (%)	RMSE	MAE	R^2_{OS} (%)
Benchmark model	1.041	0.767	–	0.610	0.461	–	0.298	0.221	–
Mean_Senti model	1.016*	0.749*	4.72	0.608	0.467	0.62	0.339	0.267	−29.53
Max_Senti model	1.037	0.762**	0.83	0.614	0.465	−1.48	0.321	0.244	−15.91
Min_Senti model	1.011*	0.744*	5.57	0.606	0.466	1.25	0.346	0.274	−34.88
Avg_Senti model	1.030**	0.754*	2.04	0.609	0.464	0.07	0.329	0.254	−21.69

Note: This table reports the results of out-of-sample prediction during the financial crisis, including the RMSE, MAE, and out-of-sample R^2. The benchmark model represents that the LSTM model includes volatility history information only, whereas the sentiment model denotes that the corresponding sentiment index is added into the benchmark model.

Abbreviations: MAE, mean absolute error; RMSE, root mean squared error.

*Represents statistical significance at 1% level based on the D–M test.

**Represents statistical significance at 5% level based on the D–M test.

TABLE 6 Out-of-sample predictive accuracy with trading volume

Horizon	$\tau_t = 5$			$\tau_t = 10$			$\tau_t = 22$		
Error accuracy	RMSE	MAE	R^2_{OS} (%)	RMSE	MAE	R^2_{OS} (%)	RMSE	MAE	R^2_{OS} (%)
Benchmark model	0.346	0.239	–	0.198	0.142	–	0.123	0.083	–
Mean_Senti model	0.334*	0.222*	6.57	0.171*	0.115*	25.32	0.113*	0.076*	15.38
Max_Senti model	0.336*	0.226*	5.74	0.174*	0.121*	22.27	0.114*	0.077*	13.79
Min_Senti model	0.336**	0.225*	5.42	0.173*	0.116*	23.55	0.114*	0.076*	13.75
Avg_Senti model	0.334*	0.222*	6.91	0.172*	0.117*	24.39	0.114*	0.076*	14.89

Note: This table reports the results of out-of-sample prediction, including the RMSE, MAE, and out-of-sample R^2. The benchmark model represents that volatility history information and trading volume as explanatory variables are included in the LSTM model, whereas the sentiment model denotes that the corresponding sentiment index is added into this benchmark model.

Abbreviations: MAE, mean absolute error; RMSE, root mean squared error.

*Represents statistical significance at 1% level based on the D–M test.

**Represents statistical significance at 5% level based on the D–M test.

required. To address this, the investor needs to estimate the mean and variance of risk premiums. For mean forecasts, we exploit a 3-year historical average (HA) to estimate the excess return in the next day. This is a benchmark method, and it is difficult to beat (Campbell & Thompson, 2008). For variance forecasts, the standard LSTM model (history information only) and sentiment models described above are used.

First, we consider using the benchmark model (LSTM) to forecast the stock volatility. In time t, the mean–variance investor will allocate the following weight of his/her portfolio to stock in period $t+1$:

$$w_{b,t} = \bar{r}_{t+1} / \left(\gamma \cdot \hat{\sigma}^2_{t+1}\right), \tag{17}$$

where $\bar{r}_{t+1}$ refers to 3-year HA of excess return and $\hat{\sigma}^2_{t+1}$ denotes its out-of-sample predictive variance based on the LSTM model. The γ is the risk aversion coefficient, and we set it be 3, 6, and 9 for a robustness check. We then obtain the portfolio return:

$$R_{b,t+1} = w_{b,t} \cdot \bar{r}_{t+1} + r_{t+1,f}, \tag{18}$$

where $r_{t+1,f}$ is the risk-free return. Thus, the CER can be expressed by

$$\widehat{CER}_b = \hat{\mu}_b - \gamma \cdot \hat{\sigma}^2_b / 2, \tag{19}$$

where the $\hat{\mu}_b$ and $\hat{\sigma}_b$ are the mean and variance of portfolio return over the out-of-sample period, respectively.

Next, we consider using the sentiment model to forecast the stock volatility. Similarly, the stock weight of portfolio in period $t+1$ is given by

$$w_{s,t} = \bar{r}_{t+1} / \left(\gamma \cdot \tilde{\sigma}^2_{t+1}\right), \tag{20}$$

where $\tilde{\sigma}^2_{t+1}$ represents the out-of-sample predicted variance of excess return based on the sentiment model. Thus, the CER is calculated by

TABLE 7 Realized utility gains relative to the benchmark model

	Δ (γ = 3, %)			Δ (γ = 6, %)			Δ (γ = 9, %)		
Sentiment model	**Weekly**	**2-week**	**Monthly**	**Weekly**	**2-week**	**Monthly**	**Weekly**	**2-week**	**Monthly**
Mean_Senti model	0.085	0.347	0.157	0.043	0.174	0.079	0.028	0.116	0.052
Max_Senti model	−0.108	−0.247	−0.099	−0.054	−0.124	−0.050	−0.036	−0.083	−0.033
Min_Senti model	0.225	0.341	0.180	0.113	0.170	0.090	0.075	0.114	0.060
Avg_Senti model	−0.097	−0.162	−0.105	−0.048	−0.081	−0.052	−0.032	−0.054	−0.035

Note: This table reports the annual utility gains (Δ) among the sentiment models relative to the benchmark model. Three volatility horizons, that is, weekly, 2-week, and monthly, are investigated. The γ represents the risk aversion coefficient and is set to be 3, 6, and 9 for a robustness check.

$$\widehat{CER}_s = \tilde{\mu}_s - \gamma \cdot \tilde{\sigma}_s^2/2. \tag{21}$$

We choose the CSI 300 stock index and 3-month treasury bond for this exercise. In keeping with tradition, we constrain $0 \leq w_{b/s,t} \leq 1.5$. The utility gain is the difference of $\widehat{CER}_s$ and $\widehat{CER}_b$; we multiply it by 25,000 and so it is interpreted as the annual percentage. The results are shown in Table 7. We find that the mean and minimum sentiment models have a larger CER, whereas the other sentiment models have a lower CER relative to the benchmark model. This finding implies that using the mean and minimum sentiment indices to predict stock volatility is meaningful from an economic point of view. On the whole, the minimum sentiment performs better than the mean sentiment index in terms of CER gains, which suggests that the consideration of sentiment memory makes economic sense. In addition, we find that the CER gains become lower with the γ higher. This is consistent with the findings of Wang et al. (2016). Although the maximum and average sentiment indices perform better than the benchmark model in data analysis as reported in Table 4, the economic value seems to be poor.

5 | CONCLUSION AND DISCUSSION

In this paper, we investigate the impacts of investor sentiment, measured by textual sentiment extracted from an investors' community, on the Chinese market stock volatility prediction. Firstly, by using the Emoji dictionary, trending word dictionary, DUT dictionary, and L&M dictionary, an integrated sentiment dictionary in the financial field is proposed for the text analysis task. We then construct four sentiment indicators to measure investor sentiment based on the dictionary-based method. When modeling the sentiment indices, we take the memory persistence of investors into account, which is proved to be effective and meaningful from an economic point of view in the empirical analysis.

Secondly, we analyze the linear relationship, including correlation and causality, between stock volatility and sentiment indices. Correlation analysis shows that all of the sentiment indices have a significant negative correlation with stock volatility. Especially, this correlation gets stronger during the financial crisis. Besides, the Granger causality analysis tells that stock volatility is the cause of sentiment indices but not the effect.

Finally, we show that sentiment indices have an impact on stock volatility based on the nonlinear causality analysis. Furthermore, we study the nonlinear predictability of sentiment indices on stock volatility by using the LSTM model. The empirical results indicate that the sentiment model achieves better performance relative to the benchmark model, which uses historical information only. The predictive accuracy increases 3.67% in terms of RMSE (the mean sentiment model when $\tau_t = 10$), 5.88% in terms of MAE (the minimum sentiment model when $\tau_t = 10$), and 7.19% out-of-sample R^2 (the mean sentiment model when $\tau_t = 10$), respectively. The fitness figures present that the work of sentiment indices are reflected in the "turning point." That is to say, sentiment helps to capture the sudden changes in the direction of the volatility. Besides, the empirical result shows that sentiment has significant predictability on short-horizon volatility rather than long-horizon volatility during the financial crisis. This because sentiment changes faster and unpredictably in this period while the long-term volatility varies relatively slow because it contains more historical information. In addition, predictive power is still significant after controlling the trading volume variable. As a result, investor sentiment measured by textual sentiment could be an effective explanatory variable in predicting Chinese stock market volatility. As a portfolio application, we explain that using textual sentiment to forecast stock volatility is economically meaningful.

This paper aims to investigate the predictive power of investor sentiment measured by textual sentiment on stock volatility. By using both linear and non-linear frameworks, we show that the non-linear model is an appropriate tool to realize this task. However, two points need to be discussed. The first discussion is about the text analysis method. The dictionary-based method is used in this study. It is difficult to verify the accuracy of the sentiment classification even though we try to make the sentiment dictionary more comprehensive and financial. For further research, one may attempt to combine ML techniques with dictionary-based methods for sentiment analysis. The second discussion is about data size. We use deep learning techniques to investigate the impact of sentiment index, and our samples include 1230 observations. Bigger data size may be more suitable for the deep learning task. Therefore, one may also expand the time interval, which represents more data, and need to download more textual documents.

ACKNOWLEDGMENT

We would like to thank Derek W. Bunn (the editor), Murali, Sowmya (the administrator), Sharmane T. Armintia (the production editor), and two anonymous reviewers for improving the quality of this paper a lot. And of course, the authors are responsible for the consequences of this article. This work was supported by the National Natural Science Foundation of China (Grant Nos. 71720107002 and 71771091), the Foundation for Key Program of Ministry of Science and Technology of China (Grant No. 2020AAA0108404), and the Financial Service Innovation and Risk Management Research Base of Guangzhou.

DATA AVAILABILITY STATEMENT

Data available on request from the authors

ORCID

Xue Gong https://orcid.org/0000-0002-0599-4740

REFERENCES

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, *57*(3), 1047–1091.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*(4), 885–905.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259–1294.

Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, *36*, 334–357.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, *61*(4), 1645–1680.

Behrendt, S., & Schmidt, A. (2018). The twitter myth revisited: Intraday investor sentiment, twitter activity and individual-level stock return volatility. *Journal of Banking and Finance*, *96*, 355–367.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Brooks, C., & Persand, G. (2003). Volatility forecasting for risk management. *Journal of Forecasting*, *22*(1), 1–22.

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, *11*(1), 1–27.

Bucci, A. (2020). Cholesky–ann models for predicting multivariate realized volatility. *Journal of Forecasting*, *39*(6), 865–876.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, *21*(4), 1509–1531.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management Science*, *53*(9), 1375–1388.

Deeney, P., Cummins, M., Dowling, M., & Bermingham, A. (2015). Sentiment in oil markets. *International Review of Financial Analysis*, *39*, 179–185.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business and economic statistics*, *20*(1), 134–144.

Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669.

Gospodinov, N., Gavala, A., & Jiang, D. (2006). Forecasting volatility. *Journal of Forecasting*, *25*(6), 381–400.

Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and control*, *2*, 329–352.

Guégan, D., & Renault, T. (2020). Does investor sentiment on social media provide robust information for bitcoin returns predictability? *Finance Research Letters*, *38*, 101494.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, *124*, 226–251.

Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies*, *28*(3), 791–837.

Huang, J.-Y., & Liu, J.-H. (2020). Using social media mining technology to improve stock price forecast accuracy. *Journal of Forecasting*, *39*(1), 104–116.

Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, *132*(1), 126–149.

Johnman, M., Vanstone, B. J., & Gepp, A. (2018). Predicting FTSE 100 returns and volatility using sentiment analysis. *Accounting and Finance*, *58*, 253–274.

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, *33*, 171–185.

Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint*. https://arxiv.org/abs/1412.6980

Lee, W. Y., Jiang, C. X., & Indro, D. C. (2002). Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking and Finance*, *26*(12), 2277–2299.

Liu, Y. (2019). Novel volatility forecasting using deep learning–long short term memory recurrent neural networks. *Expert Systems with Applications*, *132*, 99–109.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, *66*(1), 35–65.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230.

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, *42*(24), 9603–9611.

Olah, C. (2015). Understanding lstm networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs

Ren, R., Wu, D. D., & Liu, T. (2018). Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Systems Journal*, *13*(1), 760–770.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, *104*(2), 288–302.

Uygur, U., & Taş, O. (2014). The impacts of investor sentiment on returns and conditional volatility of international stock markets. *Quality and Quantity*, *48*(3), 1165–1179.

Wang, W. (2018). The mean–variance relation and the role of institutional investor sentiment. *Economics Letters*, *168*, 61–64.

Wang, Y., Ma, F., Wei, Y., & Wu, C. (2016). Forecasting realized volatility in a changing world: A dynamic model averaging approach. *Journal of Banking and Finance*, *64*, 136–149.

Wang, Y.-H., Keswani, A., & Taylor, S. J. (2006). The relationships between sentiment, returns and volatility. *International Journal of Forecasting*, *22*(1), 109–123.

Xu, L., Lin, H., Pan, Y., Ren, H., & Chen, J. (2008). Constructing the affective lexicon ontology. *Journal of the China Society for Scientific and Technical Information*, *27*(2), 180–185 (in Chinese).

Yu, J., & Yuan, Y. (2011). Investor sentiment and the mean–variance relation. *Journal of Financial Economics*, *100*(2), 367–381.

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, *55*(4), 919–926.

Zhang, S., Wei, Z., Wang, Y., & Liao, T. (2018). Sentiment analysis of chinese micro-blog text based on extended sentiment dictionary. *Future Generation Computer Systems*, *81*, 395–403.

Zhang, X., Shi, J., Wang, D., & Fang, B. (2018). Exploiting investors social network for stock prediction in china's market. *Journal of Computational Science*, *28*, 294–303.

Zhou, G. (2018). Measuring investor sentiment. *Annual Review of Financial Economics*, *10*, 239–259.

Zoen Git Hiew, J., Huang, X., Mou, H., Li, D., Wu, Q., & Xu, Y. (2019). Bert-based financial sentiment index and LSTM-based stock return predictability. arXiv preprint arXiv:1906.09024.

AUTHOR BIOGRAPHIES

Weiguo Zhang is a full research professor of decision science at the School of Business Administration, South China University of Technology, Guangzhou, China. He received his PhD degree in Financial Mathematics (2003) from Xi'an Jiaotong University, Xi'an, China. And he received his BA and MA in Math from Ningxia University and Sichuan University, respectively. His main research interests include decision theory and methodology, financial engineering, investment and risk management, and information scientific computing. Prof. Zhang has published in top-tiered journals, including *IEEE TC*, *IEEE TFS*, *EJOR*, *INS*, *IME*, *Automatica*, *IJPE*, *IJPR*, *OR Spectrum*, *FSS*, *ANOR*, *KBS*, *ASC*, *AE*, and *IREF*.

Xue Gong is currently a PhD candidate in Management Science and Engineering at the School of Business Administration, South China University of Technology, Guangzhou, China. He received his bachelor's degree (2016) from the School of Science, Yanbian University. He is very interested in the research field of behavioral finance, time series analysis, machine learning, and risk management. More information can refer to his ORCID.

Chao Wang received his bachelor's degree (2014) from the School of Mathematics and Computer, Ningxia University, China, and PhD degree in management science and engineering (2020) from South China University of Technology. His research mainly focuses on text analysis, deep learning, and credit risk evaluation.

Xin Ye received her bachelor's degree (2019) in applied mathematics from South China University of Technology. Currently, she is a PhD student in Statistics at the Wuhan University of China. Her research interests mainly include high-dimensional data analysis, model selection, and machine learning.

How to cite this article: Zhang, W., Gong, X., Wang, C., & Ye, X. (2021). Predicting stock market volatility based on textual sentiment: A nonlinear analysis. *Journal of Forecasting*, *40*(8), 1479–1500. https://doi.org/10.1002/for.2777