Analyzing Stock Market Trends Using Social Media User Moods and Social Influence

Daifeng Li
School of Information Management, Sun Yat-sen University, Guangzhou 510006, China. E-mail: lidaifeng@mail.sysu.edu.cn

Yintian Wang
School of Economics and Management, Tsinghua University, Beijing 100084, China. E-mail: wangyt2@sem.tsinghua.edu.cn

Andrew Madden
School of Information Management, Sun Yat-sen University, Guangzhou 510006, China. E-mail: admadden@hotmail.com

Ying Ding
School of Informatics & Computing, Indiana University, Bloomington, IN 47405 and School of Information Management, Wuhan University, Wuhan, Hubei 430072, China and University Library, Tongji University, Shanghai 200092, China. E-mail: dingying@indiana.edu

Jie Tang
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China. E-mail: jery.tang@gmail.com

Gordon Guozheng Sun
Technology and Engineering Group, Tencent Company, Beijing 100080, China. E-mail: gordon.gzsun@gmail.com

Ning Zhang
School of Information Management, Sun Yat-sen University, Guangzhou 510006, China. E-mail: zhangn78@mail.sysu.edu.cn

Enguo Zhou
School of Information Management, Sun Yat-sen University, Guangzhou 510006, China. E-mail: 13247695480@126.com

Information from microblogs is gaining increasing attention from researchers interested in analyzing fluctuations in stock markets. Behavioral financial theory draws on social psychology to explain some of the irrational behaviors associated with financial decisions to help explain some of the fluctuations. In this study we argue that social media users who demonstrate an interest in finance can offer insights into ways in which irrational behaviors may affect a stock market. To test this, we analyzed all the data collected over a 3-month period in 2011 from Tencent Weibo (one of the largest microblogging websites in China). We designed a social influence (SI)-based Tencent finance-related moods model to simulate investors' irrational behaviors, and designed a Tencent Moods-based Stock Trend Analysis (TM_STA) model to detect correlations between Tencent moods and the Hushen-300 index (one of the most important

Additional Supporting Information may be found in the online version of this article.

Received August 14, 2017; revised October 28, 2018; accepted November 7, 2018

 DOI: 10.1002/asi.24173

financial indexes in China). Experimental results show that the proposed method can help explain the data fluctuation. The findings support the existing behavioral financial theory, and can help to understand short-term rises and falls in a stock market. We use behavioral financial theory to further explain our findings, and to propose a trading model to verify the proposed model.

Introduction

Microblogging sites such as Twitter (https://twitter.com/) and Sina Weibo (http://weibo.com/) allow subscribers to create, share, and comment on messages of a specified maximum length (in both Twitter and Sina Weibo, the maximum is 140 characters). The format allows users to respond spontaneously and rapidly to topics, often shaping opinions. Methods of responding include forward, mention, and reply functions, which are similar to Twitter. The sites also provide comment and mail functions for users to communicate. Twitter, the world's largest microblog, has been studied widely, which provides valuable information about user sentiment. In China, the most representative microblogging websites are Sina Weibo and Tencent Weibo, each of which contains more than 0.4 billion registered users. The term Weibo means microblog. Just as messages in Twitter are called tweets, messages in Weibo are referred to as weibos.

The Bank of England's chief economist, Andy Haldane, was recently quoted as saying that data sources that provide insights into mood and sentiment can help to understand decision-making behaviors, which rarely conform to the rational processes assumed in many economic models. Developments in sentiment analysis have allowed some researchers to explore ways to use Twitter content to predict trends in stock markets. Studies have found that it can be used to improve prediction of the Dow Jones Index (Bollen, Mao, & Zheng, 2011; Pagolu et al, 2016); the NASDAQ (Bordino et al., 2012; Corea & Cervellati, 2015); and S&P 500 (Mao, Wang, Wei, & Liu, 2012). More importantly, developments in computer science have made it possible to explore micro-level economic mechanisms in greater detail, allowing the analysis of high-frequency time series (at resolutions of seconds or even microseconds) (Andersen, Bollerslev, & Cai, 2000; Harris, 1986; Wood, McInish, & Ord, 1985). Such developments can help increase an understanding of stock markets and improve the efficiency of capital allocation by computer-based trading systems. About 10–40% trading is estimated to be computer-generated high-frequency trading (Aldridge & Krawciw, 2017). Programs often buy or sell based on inferences of mood regarding the market, so a better understanding of investors' short-term moods towards stock markets is an important research direction that may improve the accuracy and efficiency of trading, and reduce the risks associated with uncontrolled high-speed trading (Aldridge & Krawciw, 2017; Carol, 2012).

One challenge for research is to explore unexplained fluctuations in high-frequency time series. Classic financial theory (Efficient Market Hypothesis) assumes that participants in a market make rational choices, and so cannot account for the unexplained fluctuations that arise because of irrational behavior (Nofsinger, 2005). Such behavior can significantly affect prices and returns. Behavioral finance rethinks market choices by modeling the public's behavior at a micro-level. The theory considers that in high-frequency time series, the investors cannot collect enough information during a short time period to help them make rational decisions. It incorporates theories from cognitive psychology (Nofsinger, 2005; Scharfstein & Stein, 1990). Researchers also attempt to use the theory of behavioral finance to explain unexplained fluctuations in stock markets (Barberis et al, 2002; Lee, Jiang, & Indro, 2002; Brown et al, 2004). However, this research does not always take new data sources (such as microblogging data) into consideration. Although some researchers have found significant correlations between social media moods and stock markets, they have not always attempted to explore the correlations in the light of current theory. The research reported in this article seeks to examine significant correlations that emerged, through the lens of behavioral financial theory.

In behavioral financial theory, investors' moods or opinions are important indicators, and can determine behaviors in stock markets. The moods of investors are often influenced by others' opinions, resulting in a "herd" effect. This is referred to as "contagion" in behavioral financial theory. According to our descriptive statistical analysis (section Experimental Results), Tencent users who use words related to finance to express their opinions are likely to be potential investors. Consequently, the modeling of moods expressed by these Tencent users' and the influence of their moods on other users' actions is based on behavioral financial theory, and provides an effective explanation of our findings.

Although there has been a lot of research into the correlation between social media moods and stock prices, few studies have focused on high-frequency time series, especially time-series relating to China's stock market. Our main contributions are summarized below:

- We introduce an algorithm that combines sentiment analysis with SI to model two important aspects of financial behaviors: investor sentiment and mood diffusion. Under the direction of behavioral financial theory, SI is mainly used to simulate the diffusion of influential opinions and its effect on user moods across the social network. More important, SI can help to resolve the problem caused by the disparity in frequency between weibo data and Hushen-300 time series data. The latter are collected at 5-minute intervals, and there is not always sufficient weibo data in a given 5-minute interval to infer a mood.
- We incorporate GARCH with VAR regression (TM_STA) to detect correlation patterns between SI-based Tencent moods and the Hushen-300 indicator. This helps to verify that the proposed model can reflect investors' irrational behaviors in the stock market, and can help to explain unexplained fluctuations.
- We use three typical financial behavior theories, the Sheep-Flock effect (Scharfstein & Stein, 1990), Positive feedback effect (De Long, Andrei, Sumlners, & Waldmann, 1990; Hirshleifer, Subrahmanyam, & Titman, 2002) and Optimistic Expectation

DOI: 10.1002/asi

theory (Blandchard & Woston, 1982) to explain our findings based on the proposed methodology. New experiments have been added to help test the possibility that each of the above theories is having an impact.

This article presents both theoretical and methodological innovations. At a theoretical level, we present a novel analysis of a data set, and use it to model financial behaviors. The model derived supports previous studies of behavioral finance, and identifies new patterns, which emerge because of the use of high-frequency time series. At a methodological level, we introduce a new model TM_STA to handle the unexplained fluctuation in high-frequency time series. These methodological innovations have practical values for future research in related areas.

Related Work

Sentiment Analysis

One key objective of sentiment analysis is to gain an indication of feelings on a particular topic. Various efforts have been made to mine the knowledge and experience assumed to be present in cyberspace, and to use available data to gain insights into factors that shape public opinion (Li et al, 2012; Patel, Prabhu, & Bhowmick, 2015; Ravi & Ravi, 2015; Bollegala et al, 2016). Nguyen, Wu, Chan, Wei, and Zhang (2012) developed models based on time-series to predict opinion from Twitter data. Wang, Tong, and Chin (2014) used enhancement technologies to improve the efficacy and accuracy with which sentiments inferred from Twitter data are classified. Devi, Palaniappan, and Kumar (2015) applied an ensemble of machine learning approaches to realize feature selection for tweet sentiment prediction. Tang, Qin, and Liu (2015) introduced Conv-GRNN and LSTM-GRNN for document level sentiment classification. Zhou et al. (2016) generated syntax trees of sentences, then introduced multiple sentiment features into the traditional basic words-bag features. There also exists a lot of research that has focused on social media in Chinese, including microblogs such as Sina Weibo (Chen, Bai, & Zhan, 2014; Zhang, Chen, Liu, & Wang, 2017) and social networking sites, such as Douban (https://www.douban.com/) (Yang & Yecies, 2016).

In behavioral financial theory, investor sentiment is also an important research direction. Many researchers have used the behavior of stock markets to infer investors' moods (Lee, 1991; Neal & Wheatley, 1998). Based on the assumption that mood will influence stock market trends, they have tried to explain fluctuations which cannot be well explained by using classic finance theory (Brown, 1999). By using Tencent weibos containing specialist financial terminology (for example, bullish/bearish) we have been able to model users' moods towards the stock market more directly. Although similar studies have been made using microblog data, this is the first time that a model has been developed that uses high frequency time series, and attempts to take social influence into consideration.

Social Influence

In social networks, using related postings provides insights into SIs and their impact on users. For example, Tang, Sun, Wang, and Yang (2009), and Liu, Tang, Han, Jiang, and Yang (2010) used different text-mining approaches to assess social influence in data networks, but both found it to have predictive value. Tan et al. (2011) analyzed the social networks of Twitter users and found that their approach led to significant improvements in sentiment analysis. Li et al. (2012) found that a model, which linked the topic influence of Weibos to sentiments expressed, could help predict users' opinions. Social influence is an efficient method to use network information to predict users' behaviors, especially for solving the data sparsity problem.

In another aspect, human interaction is known to play a part in the value of the stock market, with some people's propensity for interaction acting like a "social multiplier" (Nofsinger, 2005). This multiplier effect is explored further by DeMarzo, Vayanos, and Zwiebel (2003), who analyzed a range of social factors that affect people's interactions with financial markets. In behavioral financial theory, research considers an investor's behavior will be significantly influenced by crowds (or herds). People imitate other people, making behavior contagious (Scharfstein & Stein, 1990). Our model incorporates social influence and simulates its influence on user moods.

Behavior Finance Model Based on Social Media

According to behavioral finance theory (Bauman, 1967; Burrell, 1951; Slovic, 1972), the trends of finance indicators do not fully satisfy a random walk, yet they can be predicted to some extent by analyzing investors' irrational behaviors; for example, the Sheep-Flock Effect (Scharfstein & Stein, 1990), especially in relation to investor sentiment, which is considered a systematic factor.

In the finance domain, behavior modeling is mainly based on financial indicators, such as the Pull-Call Ratio (Dennis et al., 2002), Barron's Confidence Index (Lashgari et al, 2000), Closed-end Fund Discount (Lee et al, 1991), and so on. To handle high-frequency situations, researchers used an investors' utility function based on a theory of value function (Tversky et al, 1992) to model investors' confidence and risk aversion. A number of studies have confirmed that the theory of behavioral finances gives a better explanation of unexplained phenomena; for example, incomplete and unbalanced information distribution may cause concentrated trading; and investors' sentiments are strongly correlated with unexplained fluctuations in the stock market (Lee et al., 2002).

Recent research clearly supports the view that investor sentiments in social media can influence financial markets. Many researchers considered that an investor's decision-making is influenced by social moods; and stock market trends can help forecast future financial and economic

DOI: 10.1002/asi

TABLE 1. Summarization of Tencent Weibo from Oct. to Dec.

Items	Users	Original	Forwards	Replies	Comments	Mails	Mentions
Total Number	326,497,021	3,607,924,594	1,026,243,542	43,658,112	299,354,146	174,440,376	2,347,927

activity (Antweiler & Frank, 2004; Da, Engelberg, & Gap, 2015; Nofsinger, 2005; Telock, 2007; Borzykowsi, 2011; Yang & Zhou, 2015). Some recent research also considered the correlation between social influence and the stock market (Benthaus & Beck, 2015; Nofer & Hinz, 2015). Some researchers studied social media in China, such as Sina Weibo, and their influence on the stock market (Chen, Cai, & Lai, 2016). Research of social moods and its influence over fluctuations in stock price has also been taken into consideration, for example, by Bakshi et al (2016) and Karabulut (2013). Unlike previous studies, we mainly consider building a correlation between social media and the stock market from the perspective of high-frequency time series, the main challenge of which is the unexplained fluctuation caused by irrational decisions.

Methods

Experimental Data

Data from Tencent Weibo and the Hushen-300 Index were analyzed. The data were collected between October 1 and December 31. A summary of the Tencent data is provided in Table 1. For the data of Hushen-300, we first divided them into the morning trading time (9:20 AM to 11:30 AM) and afternoon trading time (13:20 PM to 15:30 PM). We collected the price of the Hushen-300 Index every 5 minutes and excluded weekends and vacations. After that, we obtained 2,486 observations of Hushen-300. Then we sorted all Tencent data according to their time stamps, divided them into subgroups based on the time division of Hushen-300 (every 5 minutes during the morning and afternoon trade times, excluding weekends as a subgroup). We also considered before trading time and noon time, and lastly, we obtained 2,542 subgroups, where for the *tth* subgroup, we could get the entire sentiments at time stamp *t*.

Framework of the Proposed Model

The framework of the proposed model is shown in Figure 1. We introduce a new approach to modeling financial behavior based on sentiment analysis and social influence inferred from social media. Data from Tencent Weibo were processed to generate an SI-based Tencent mood time series developed with reference to behavioral finance theory. Following this, we used TM_STA, a model based on a combination of VAR and GARCH models, to analyze correlations between Tencent moods and the Hushen-300 time series. This helps to illustrate the effectiveness of the

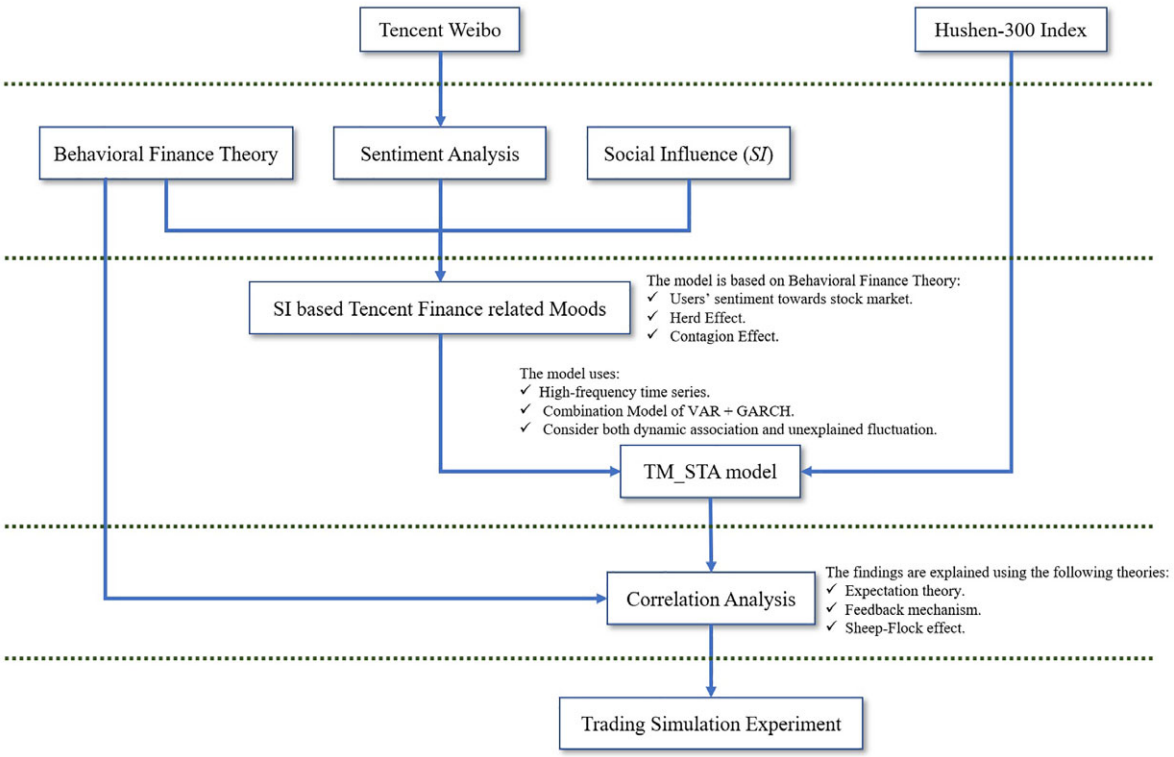

FIG. 1. The framework of the proposed model. [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/asi

TABLE 2. Notation.

Variable name	Description
ES	ES includes the 22 most representative finance-related entities, which reflect the stock market in China. Examples include: finance market, stock, Blue-chip, A shares, B shares, CSI-300, and so on. The entities are mainly from two websites: **SOGOU:** https://pinyin.sogou.com/dict/detail/index/15127?rf=dictindex **BAIDU:** https://baike.baidu.com/item/%E8%82%A1%E7%A5%A8%E6%9C%AF%E8%AF%AD/4642854?fr=aladdin **Limits/Boundaries:** Most representative entities related with stock market.
DS	A dictionary of sentiments, which includes positive and negative sentiment vocabulary. The dictionary is mainly compiled from Chinese Academy of Social Science (General Sentiment words): http://ling.cass.cn/pbfiles/yys_jj_e.asp We also added 71 domain-negative sentiment words, and 26 domain-positive sentiment words into our dictionary. Domain-positive and domain-negative sentiment words are specialist terms highly related with financial domain, such as "be Bullish" or "be Bearish" (Finance-related sentiment words). **Limits/Boundaries:** General sentiment words, Finance-related sentiment words
IFB	IFB (Indirect Forwarding behavior): IFB means if user *A* forwards *B* on Topic *T*, *C* forwards *A* on Topic *T*, then we consider that *C* has a high probability to agree with *B*'s opinion on Topic *T*. **Limits/Boundaries:** We only consider two-stage indirect forwarding behaviors.
TM_STA	TM_STA incorporates both VAR and GARCH into a unified framework to handle with high-frequency situation. **Limits/Boundaries:** For VAR, we make Logarithmic and Difference for each variable to obtain stable time series. For GARCH, we use the common assumption that the variance σ_t^2 is highly influenced by σ_{t-1}^2.
TencentMood_Num{*t*}	*TencentMood_Num*{*t*} determines the total number of users' emotions towards identified entities in ES using a sentiment detection algorithm (it includes both positive and negative sentiments during time period t-1 and t). **Limits/Boundaries:** The time range is from 2011.10.01 to 2011.12.31.
TencentMood_Percent{*t*}	*TencentMood_Percent*{*t*} determines the percentage of *TencentMood_Num*{*t*}. **Limits/Boundaries:** The time range is from 2011.10.01 to 2011.12.31.
Hushen − 300{*t*}	The data relating to the Hushen-300 indicator is derived from the authoritative finance website http://www.wind.com.cn/NewSite/edb.html **Limits/Boundaries:** The time range is from 2011.10.01 to 2011.12.31.

proposed model in explaining some of the previously unexplained fluctuation apparent in high-frequency time series. Three typical financial behavior theories were used to explain our findings, and a trading simulation was run in order to validate the effectiveness of our proposed model. We constructed Table 2 to describe each variable in detail.

Sentiment Analysis on Tencent Weibo

For the purpose of sentiment analysis relating to the stock market, three subtasks must be performed. It must be determined, Firstly whether the current microblog is related to any economic trend; secondly, whether the current microblog reflects subjective sentiment; and thirdly, how to identify how the user feels about the economic trend.

For the first subtask, we used a corpus *ES* comprising 22 domain entities (see Table 2), to select microblogs relating to economic issues. For the second subtask, we applied several constraints and assumptions with a view to distinguishing between subjective and objective statements, according to methods proposed by Yu and Hatzivassloglou (2003). For the third subtask, we used FudanNLP (a Chinese NLP tool) to carry out grammar dependency analysis and to detect users' opinions of the current entity (that is, the subject of discussion in the current microblog). This analysis was based on the dictionary of sentiments *DS* (see Table 2). In circumstances where FudanNLP is invalid (such as informal sentences) PMI (Turney, 2002) was applied and significantly improved the recall of the sentiment detection algorithm.

To achieve the third task, we relied on *DS*, an extended version of a dictionary of sentiments compiled by the Chinese Academy of Social Science.[1] This dictionary contains 14 different mood classifications, including: Happy, Angry, Sad, Loving, Hating, Worrying, Fearful, Regretful, Surprised, Missing, Respectful, Calm, Disappointed, and Excited. In *DS*, we added 71 domain-negative sentiment words and 26 domain-positive sentiment words. Domain-positive and domain-negative sentiment words are specialist terms highly related to financial domains, such as "be Bullish" or "be Bearish." Such terms indicate subjective sentiments. We therefore consider them to be reflections of a user's sentiment at the time the weibo was posted. For example, "看空"[2] is a common Chinese response to something that is perceived to be a bear market and is an indication of pessimism. "看多"[3] is the opposite. Table 3 shows examples by using different sets of sentiment words.

Social Influence Model

Crowds' moods towards stock markets are important in financial behavior research, hence the potential value

[1] http://ling.cass.cn/pbfiles/yys_jj_e.asp
[2] Literally, "See empty."
[3] "See more."

DOI: 10.1002/asi

TABLE 3. Example of weibos using sentiment dictionary *DS* to describe the finance market.

	User ID	Weibos
Weibos with Finance-Related sentiment words.	2202	I predict the stock market will rise continuously for one month. We could buy in around 2,100.
	94391	Where is the bottom of the stock market? Should we prepare for bargain-hunting?
	991	Visiting a group of friends from several social circles - they consider the stock market is almost touching the bottom and should be bought now.
Weibos with General sentiment words.	6021	I entered the stock market in the morning, and cry in the afternoon.
	93718	Stock market, turbulence and tragedy.
	39088	Give up from stock market, weeping.

of indicators of mood such as microblogs. However, in Tencent Weibo, if users' moods are collected over very short time periods (for example, 5 minutes) the problem of data sparsity arises because the messages are not distributed evenly along the time line. Processing the messages to identify mood may not reflect the real sentiment of a finance market. To address this problem, an SI model is used to infer opinions and complement Tencent moods data. This can describe how moods are diffused and influenced, which is a key element in the herd effect associated with financial behavior. Thus, the SI model offers an effective way to fit users' herd behavior and irrational behavior at the micro-level, providing more interpretable results based on behavioral financial theory. In our model, therefore, we focus on behaviors indicative of social interaction, such as forwarding of weibos, replies, comments, mails, and other communications between two users on the same topic. Based on previous studies (Kwak, Lee, & Moon, 2010; Tan et al., 2011), we make the same assumption that, if user B has forwarded user A's weibo relating to topic T, then there is a high probability that B shares A's social mood. Although there are instances where forwarders disagree with the opinion of the original posted messages, just as discussed by previous studies, the percentage is assumed to be too small, to cause a significant negative influence.

In order to further model the diffusion process or "contagion" phenomenon in financial behavior, indirect forwarding behavior (IFB) (Dong et al., 2012; Zhuang et al., 2012) is also taken into consideration. The definition of IFB is shown in Table 2. Zhuang et al. (2012) indicate that a number of meta indirect influence structures can improve prediction performance on different social networks, such as Twitter, E-mail, and Mobile networks. We analyze patterns associated with IFB in Tencent weibo and use the analysis to generate a predictive model based on Factor Graph models (Zhuang et al., 2012). The objective function can be represented as:

$$L(ys|u_A,u_B,T)=\frac{1}{Z}\sum_{i\in X}\theta_i\times\log\mu_i(ys,u_A,u_B,T) \qquad (1)$$

where $ys \in (+1, -1)$, $+1$ means "Forward," and -1 means "Not Forward." $L(ys|\ u_A,\ u_B,\ T)$ is the log likelihood of the probability distributions of u_B for forwarding u_A on Topic T (Tang J et al, 2008). X is the set of attributes, which includes the number of "Forwards," "Replies," "Comments," "Mails," and "Indirect Structures" between user u_A and u_B. Z represents normalized factors and θ_i represents the parameters of function μ_i. The functions μ_i between users u_A and u_B satisfy exponent increase functions, which are mainly based on the statistical analysis shown in Figure 2. Some selected attributes are used to show its inner relationships with the probability of Forwards. For example, an increase in the number of replies, mails, and comments exchanged between two users is associated with a significant increase in the probability that one of the users will forward weibos of the other (the increasing curve in log–log scale; 99% instances satisfy the positive correlations). In addition, in Figure 2d, the total number of IFBs can also provide a positive contribution towards the probability of a Forward.

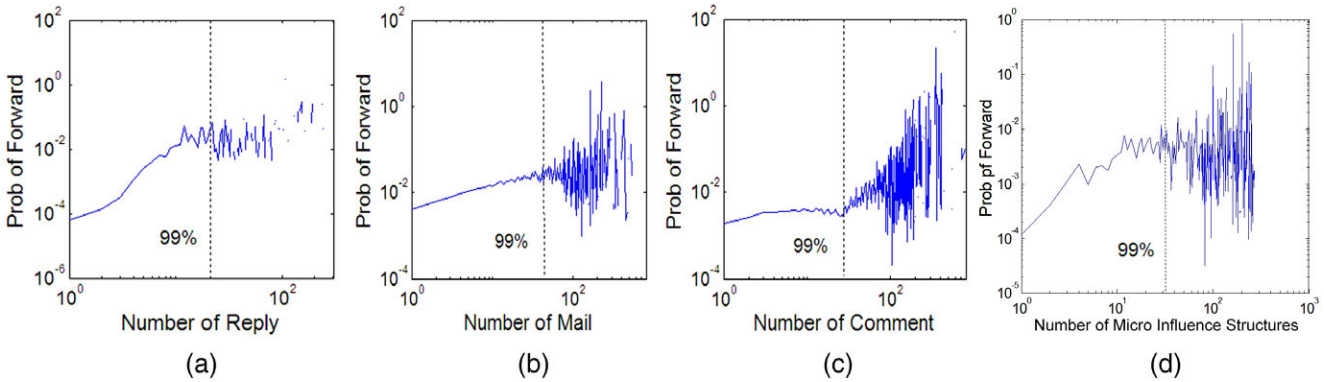

FIG. 2. "Forward" probability vs. selected attributes. [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/asi

TM_STA Model

VAR regression. Vector Auto Regression (VAR) (Simkins, 1994) is widely used to evaluate relationships between different global economic indicators. One of its main advantages is that it can fit influences, caused by different social pulses. This means that we can fix other variables, change the value of a current social variable to see what changes can happen to the whole system, and how long the pulse will last. In our research, we consider that the value of the entire Tencent mood can reflect social pulse; for example, people will be more sensitive to the finance news: when new finance news is published, a social pulse occurs. The relationship between Tencent moods and the Hushen-300 can be seen in formula 2.

$$\begin{bmatrix} DLOG(TencentMood\_Num\{t\}) \\ DLOG(TencentMood\_Percent\{t\}) \\ DLOG(Volume\{t\}) \\ DLOG(Hushen-300\{t\}) \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} + \begin{bmatrix} A^{t-d}{}_{11} & A^{t-d}{}_{12} & A^{t-d}{}_{13} & A^{t-d}{}_{14} \\ A^{t-d}{}_{21} & A^{t-d}{}_{22} & A^{t-d}{}_{23} & A^{t-d}{}_{24} \\ A^{t-d}{}_{31} & A^{t-d}{}_{32} & A^{t-d}{}_{33} & A^{t-d}{}_{34} \\ A^{t-d}{}_{41} & A^{t-d}{}_{42} & A^{t-d}{}_{43} & A^{t-d}{}_{44} \end{bmatrix} \begin{bmatrix} DLOG(TencentMood\_Num\{t-d\}) \\ DLOG(TencentMood\_Percent\{t-d\}) \\ DLOG(Volume\{t-d\}) \\ DLOG(Hushen-300\{t-d\}) \end{bmatrix} + \begin{bmatrix} E_{1,t-d} \\ E_{2,t-d} \\ E_{3,t-d} \\ E_{4,t-d} \end{bmatrix} \quad (2)$$

C is a constant and E is an error item. *DLOG* means we take Logarithmic and Difference for all selected variables, which could smooth the two time. d is time delay. TencentMood_Num{t} means calculating the total number of users' emotions, TencentMood_Percent{t} means calculating the percentage of positive sentiment at time period *t.*

Metropolis-within-Gibbs-based GARCH. To further improve understanding of the influence of Tencent moods on the Hushen-300 index, we need a model to provide the range and variance distrubution of value changes; The model can help to explain the volatility fluctuation, caused by investors' irrational decisions in a high-frequency time series. We designed the target functions based on the theory of Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) (Engle, 1982; Bollerslev, 1986; Brook & Persand, 2003), and used a Monte Carlo Markov Chain framework (Chib & Greenberg, 1995; Chib et al, 1998; Teruo, 2000) to estimate the parameters. The formula can be seen as:

$$\begin{cases} y_t = \vec{\theta} \bullet \vec{Y} + \vec{\phi} \bullet \overrightarrow{TencentMood\_Num} + \vec{\phi} \bullet \overrightarrow{TencentMood\_Percent} + \vec{\gamma} \bullet \vec{X} + \mu_t \\ \sigma_t^2 = \omega + \alpha \times \mu_{t-1}^2 + \beta \times \sigma_{t-1}^2 \end{cases} \quad (3)$$

where the error of return y_t of Hushen-300 at time t is $\mu_t \sim N(0, \omega + \alpha \times \mu_{t-1}^2)$, its variance is σ^2, ω is constant. $\vec{\theta}$ is the vector of parameters of $\vec{Y}$, which is the vector with lags for auto regression (ex: $\vec{Y} = \{y_{-1}, y_{-2}, \ldots y_{-l}\}$); $\vec{\phi}$ is the vector of parameters of $\overrightarrow{TencentMood\_Num}$ with lags for Tencent moods; $\vec{\phi}$ is the vector of parameters of $\overrightarrow{TencentMood\_Percent}$ with lags for marketing sentiment; and $\vec{\gamma}$ is the vector of parameters with lags for other variables, such as volatility and volume. The first function in Equation 3 represents a correlation among different variables, while the second one fits the abnormal fluctuation of the Hushen-300 index, which is caused by investors' irrational decisions.

Combination model. The TM_STA model combines VAR and GARCH into a unified framework. This provides more latent relationships than the use of one model only. For example, VAR can find an association pattern *PA*, while association pattern *PB* is detected only after applying GARCH (an association pattern is a correlation function). If *PA* and *PB* are combined, the resulting model may be able to handle $PA \cup PB$ situations. The formula is:

$$EX(Y_t) = \frac{P(Y_t^{VAR}) \times Y_t^{VAR} + P(Y_t^{\text{GARCH}}) \times Y_t^{\text{GARCH}}}{P(Y_t^{VAR}) + P(Y_t^{\text{GARCH}})} \quad (4)$$

where $P(Y_t^{VAR})$ means the probability of obtaining Y_t at time stamp t by applying the *VAR* model, and Y_t^{GARCH} represents the probability of obtaining Y_t at time stamp t by applying the GARCH model. $EX(Y_t)$ represents the expected Y_t by using the combined model.

Experimental Results

Statistical Analysis

Demographic analysis of Tencent users. Demographic analysis of Tencent users coupled with behavioral financial theory offers valuable insights when modeling investor sentiment. We used demographic analysis of Tencent users to address relevant questions such as "Who publishes

DOI: 10.1002/asi

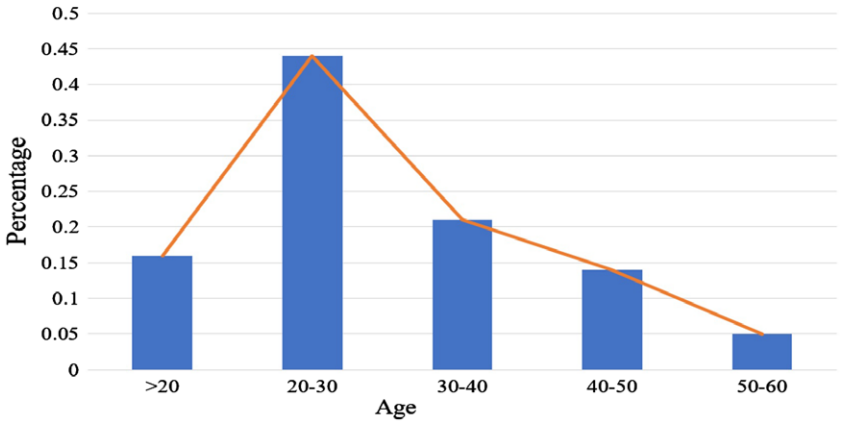

FIG. 3. Age distribution of finance-related Tencent users (1.5 million). [Color figure can be viewed at wileyonlinelibrary.com]

finance-related weibos?" and "How many of them are potential investors?" We used the entity dictionary *ES* to filter all users who mentioned finance-related topics in their microblogs. We then drew their age distribution in Figure 3.

About 1.5 million users contributed 13 million finance-related weibos to Tencent Weibo data. Among all finance-related weibos, when those who used emotional language to describe the stock market or were clearly advertising associated services were excluded, there remained about 5 million weibos that used finance-related words to describe the finance market. Examples can be seen in Table 3. The weibos were from 0.73 million users, who, it was inferred, were likely to be investors (most of them were aged between 20–50). In summary, all the 1.5 million users are potential stock investors, but 0.73 million users used finance-related language and so are more likely to be true investors. Thus, from the perspective of these descriptive statistics, Tencent data offer insights into potential investors in the stock market.

Sentiment analysis. It is important to accurately identify the moods of Tencent users towards the stock market, so we designed an evaluation experiment to verify the proposed sentiment detection algorithm. For testing data, we manually label a data set that contains 10,000 samples, which is large enough to cover almost all sentiment expression manners. The basic sentiment model, which is introduced in the Methods section, achieved high accuracy, which is 88%, but recall is around 65%. By incorporating PMI methods, both recall and accuracy reached up to 85%.

Social influence analysis. The test to assess the performance of the proposed *SI* model is shown in Table 4. We use three classical algorithms as a baseline for comparison: SVM (Support Vector Machine), LR (Logistical Regression), and CRF (Conditional Random Field). For CRF, the code is mainly from Wu et al. (2012). For SVM, we use SVMlight.[4] For LR, we adopt the Statistical Toolbox.[5]

[4] http://svmlight.joachims.org/

[5] http://www/csie.ntu.edu.tw/~cjlin/liblinear/

For a small testing data set with 11,000 target users and their weibos, the proposed social influence model SI with IFB outperforms the other baseline in predicting the Forwarding behaviors, where accuracy levels reached up to 70.11%. SI has two advantages: first, it can gain timely information relating to the influence of mood (herd effect) and diffusion (contagion effect); second, it can detect influential relationships among users with acceptable accuracy and recall, which can help resolve the issue of data sparsity. Approximately 30% of the timepoints have insufficient data. However, by supplementing sentiment analysis with social influences, user sentiments towards financial markets can be inferred with reasonable confidence.

Correlation analysis between Tencent finance-related moods and Hushen-300 index. We use positive, negative sentiment and sum(positive, negative) to perform a statistical analysis, and find that only positive sentiment has a degree of correlation with the Hushen-300 index. The possible reasons will be explained in the section "Theoretical Explanation." The results are shown in Figure 4.

Figure 4a is the time series of Positive Tencent Moods without *SI* (each time stamp represents 5 minutes); 4b is the time series of Positive Tencent Moods with *SI*; 4c is the time series of the Hushen-300 index of the same time period. Tencent Moods (both with and without *SI*) have correlations with the Hushen-300 index along the time based on direct observation. As shown by the red ellipse in 4a, there is a serious data sparsity problem. The entire

TABLE 4. Performance of forwarding predictions.

	Accuracy	Recall	F1-Score
***SI* (with IFB)**	**70.11%**	**97.53%**	**0.8158**
***SI* (without IFB)**	69.02%	96.05%	0.8088
SVM	31.88%	100%	0.4835
LR	66.09%	100%	0.7958
CRF	68.99%	96.12%	0.8088

According to previous studies (Kwak et al., 2010; Tan et al., 2011), forwarding behaviors can be considered as opinion agreement and propagation.

DOI: 10.1002/asi

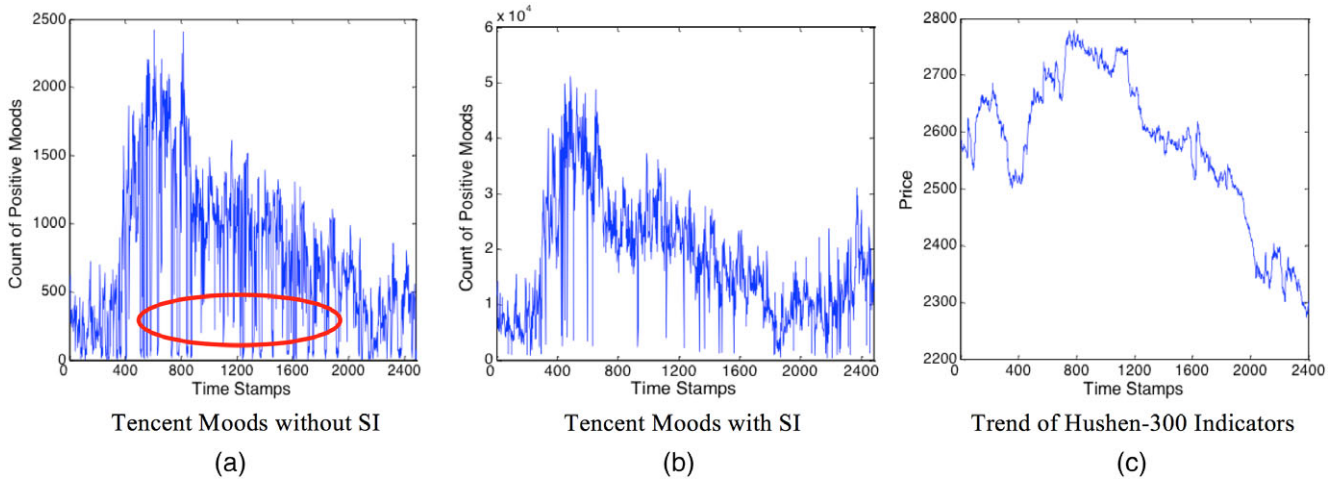

FIG. 4. Statistical analysis based on count of positive sentiment. [Color figure can be viewed at wileyonlinelibrary.com]

Tencent moods fluctuate fiercely with time. The reason is that the sentiment collection without *SI* along time is not balanced; sometimes the collection set is big (the amount may be influenced by social pulse, such as important financial news), sometimes the set is small. Figure 4b shows the result of analyzing the same data but with *SI* taken into account. This significantly reduces the noise created by lack of data, helping to smooth the curve. Trends of Tencent moods and Hushen-300 show clear similarities, which suggests a correlation between Hushen-300 and Tencent mood. But the correlation test for both with *SI* and without *SI* is not significant. We design a new variable, "Positive percentage," which is "count of positive / (count of positive + count of negative)," repeat the experiment, and find some meaningful results. Figure 5 shows the statistical analysis based on positive percentage.

Again, adjusting for *SI* helps to reduce noise, and reveals clear trends and possible correlations with the Hushen-300 data. Based on the data described in Figure 5, we made a correlation test, and summarize significant results in Table 5.

As shown in Table 5, the first number in the bracket of each cell represents the assignment of lags; for example, −9 means that Tencent (−9) is correlated with Hushen-300; the second element represents the correlation coefficient and the third represents a significant test, only when the probability is smaller than .05 is the correlation established. Tencent morning time series of moods have a significant correlation with the Hushen-300 morning time series (every 5 minutes) with lags of 9. For the afternoon data, only Tencent Moods with SI pass the significant test with lags of 6. Noon and Evening Tencent Moods with SI have significant correlations with the Hushen-300 morning and afternoon closing price. The correlation between Evening Tencent Moods and Hushen-300 afternoon closing price does not pass the significance test. Above all, compared with Tencent moods without SI, Tencent Moods with SI, has a stronger correlation with the Hushen-300 index, which indicates that SI based Tencent moods is an effective method of simulating investors' irrational behaviors in high-frequency time series.

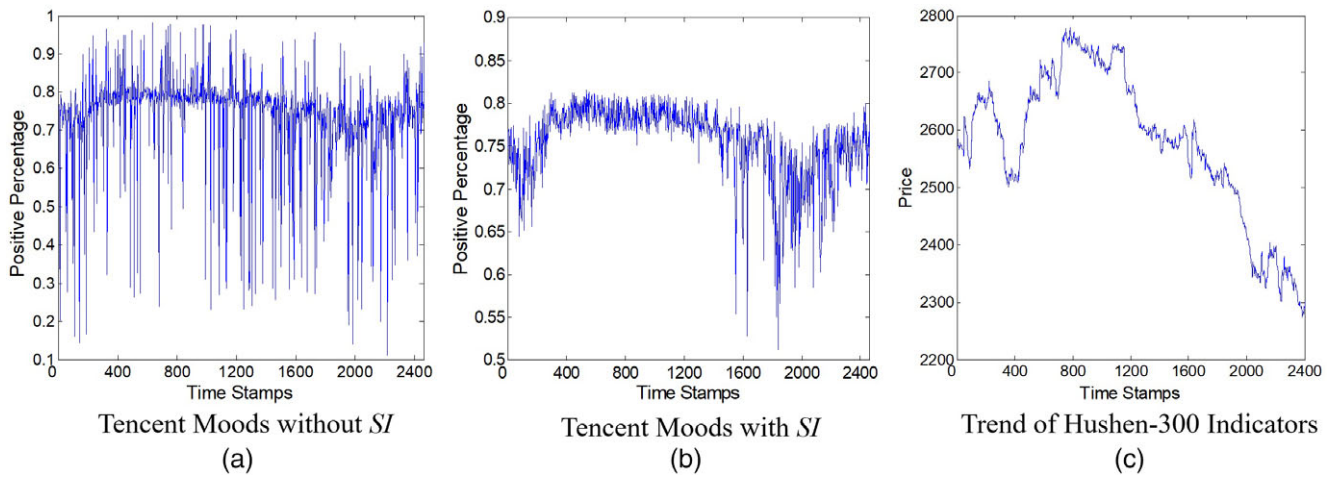

FIG. 5. Statistical analysis based on positive sentiment percentage. [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/asi

TABLE 5. Summarization of significant correlations between Tencent Moods and Hushen-300.

Tencent Moods Positive Percentage	Hushen-300 morning trend (9:20 AM-11:30 AM)	Hushen-300 afternoon trend (13:20 PM-15:30 PM)	Hushen-300 morning close price (11:30 AM)	Hushen-300 afternoon close price (15:30 PM)
Morning with *SI* (9:20 AM-11:30 AM)	(−9, 0.36, Prob <0.05) Each lag is 5 min	–	–	–
Morning without *SI* (9:20 AM-11:30 AM)	(−9, 0.32, Prob <0.05) Each lag is 5 min	–	–	–
Afternoon with *SI* (13:20 PM-15:30 PM)	–	(−6, 0.32, Prob <0.05) Each lag is 5 min	–	–
Afternoon without *SI* (13:20 PM-15:30 PM)	–	(*, *, Prob >0.05) Each lag is 5 min	–	–
Noon with *SI* (11:30 PM-13:20 PM)	–	–	(−1, 0.41, Prob <0.05) Each lag is 1 day	(−3, 0.55, Prob <0.05) Each lag is 1 day
Noon without *SI* (11:30 PM-13:20 PM)	–	–	(−1, 0.35, Prob <0.05) Each lag is 1 day	(−3, 0.52, Prob <0.05) Each lag is 1 day
Evening with *SI* (15:30 PM- Next 9:20 AM)	–	–	(−1, 0.29, Prob <0.05) Each lag is 1 day	(−1, 0.21, Prob <0.05) Each lag is 1 day
Evening without *SI* (15:30 PM- Next 9:20 AM)	–	–	(−1, 0.28, Prob <0.05) Each lag is 1 day	(*, *, Prob >0.05) Each lag is 1 day

For high-frequency situations, investors cannot collect enough information to make decisions in a short time, so the probability of irrational behaviors will be high, causing a fluctuation that cannot be explained by using traditional regression (morning and afternoon trading time). SI-based Tencent moods can better simulate this kind of situation, which is helpful in allowing GARCH to add variance, improving the fit to previously unexplained parts of the time series. However, the correlation is not significant if data for morning and afternoon trading are combined to give data for a whole day. This produces a correlation with a significance level of only .1 (so not significant at .05). Furthermore, a co-integration test of the residual sequences between two time series is also significant: when make one-order differences and assign the lag as −1 (each lag is 1 h), the coefficient is around 0.7 and $p = .029$. The test results further show that there exist steady correlations between SI-based Tencent finance-related moods and Hushen-300 time series.

Theoretical Explanation

In this section we illustrate the association of our findings with three typical financial behavior theories: the Expectation theory, feedback mechanism, and Sheep-Flock effect.

Definition 1. Expectation theory: Assume a crowd's expected probability of positive return r during time period t is p and expected probability of loss l is q (p + q = 1). The utility function of the crowds is v(t) = p * r + q * l. According to finance behaviors, if the crowd's expectation during time period t is positive, members of the crowd tend to take actions based on positive information, and ignore negative information with potential risk.

Definition 2. Feedback mechanism: Crowds' optimism regarding the market (ex. Positive sentiment) will always be taken advantage of by speculators, who will stimulate price rises by buying stock, in effect betting that some people will buy their stocks at the higher price.

Definition 3. Sheep-Flock effect: A manifestation of conformity psychology, which means that when market information is incomplete, crowds prefer to follow the opinions of influential institutions or opinion leaders.

First, we attempt to use expectation theory and feedback mechanism to explain why weibos classed as positive are often significantly correlated to price rises, while negative weibos show no significant correlation. The experimental data cover a period from 1st October 2011 to 31st December 2011. As seen in the authority report,[6] although the Hushen-300 index has continuously decreased since 2007, according to periodic theory, the latter half of 2011 was predicted to be a "bottoming-out" stage. Expectation theory suggests that the market has a strong tendency to "buy at bottom," and that crowds tend to act on positive emotions, while ignoring negative emotions. According to Feedback mechanism, if market expectation is positive, speculators will continuously act on the desire to "buy at bottom" because they consider that there will always be someone who will buy their stocks when prices rise. The resulting feedback will further increase the correlation with positive sentiments.

In addition, according to our statistical analysis, general sentiment words (example in Table 3) account for 90% of weibos expressing negative sentiment, while finance-related sentiment accounts for 85% percentage of positive sentiment (example in Table 3). The phenomenon could be another reason why only positive sentiment has a significant correlation with the stock market because users of finance-related sentiment words are more likely to be investors.

[6] http://stock.xinhua08.com/focus/2011/2012stock/

DOI: 10.1002/asi

Second, the Sheep-Flock effect will lead crowds to follow speculators and engage in similar investment behaviors: Because of high-frequency time series, crowds may not have enough time to collect enough information to make a rational decision, and their behaviors may cause irrational fluctuation, which cannot be explained by VAR regression. As described in the Statistical Analysis section, GARCH can reduce the influence of fluctuations in price volatility that cannot be explained by VAR (as shown in Table 5). The phenomenon further supports the theory that the Sheep-Flock effect contributes to volatility, especially in a high-frequency situation.

Finally, the feedback mechanism will enhance positive volatility, because with high positive expectation, investors will stimulate prices rising by buying stock, betting that some others may buy at a higher price. We designed a new experiment to further explore our findings in the context of the established theories discussed above. We used the percentage of positive sentiments at time t as an indicator of the crowd's expectation of the stock market, and we observed the fluctuations in price during $(t, t + 9)$ in the morning, and $(t, t + 6)$ in the afternoon. The experimental results are shown in Table 6.

Assuming that the sentiment percentage and price volatility at time t are $p(t)$ and $y(t)$, respectively, then Avg Dev is defined as $\left\{\sum_{t\in\{p_t>thres\}}\sum_{i=1}^{Lags}(y_{t+i}-y_t)/Lags\right\}/N(p_t>thres)$, where there is a threshold of Positive Sentiment Percentage (80%, 75%, …). Max and Min Dev are the maximum, minimum value of $\sum_{i=1}^{Lags}(y_{t+i}-y_t)/Lags$.

In Table 6, all the maximum values of Avg Dev are obtained when the percentage of positive sentiments is greater than 75%. When the percentage is lower than 75%, the Avg Dev is more likely to be negative. One possible reason is that, during the main rising stage of the price (from timestamp 380 to 1,500), almost all the positive emotion values are greater than 75%, as seen in Figure 6. The phenomenon further suggests that, if the expectation of the market is high at time t, the feedback mechanism will enhance the positive volatility over the subsequent timestamps. In addition, when the percentage of positive sentiment is high (above 75%), the overall volatility (including both positive and negative Dev) is enhanced significantly (as seen in Figure 6, where the total Dev is 2,780–2,500 = 280). This phenomenon is in accord with existing research (Lee et al., 2002; Brown et al; 2004), which indicates that the Sheep-Flock effect will increase irrational fluctuation.

Evaluation

Previous studies mainly use statistical methods to detect the correlation between social media moods and stock market, but seldom consider high-frequency time series. We use the most common method, VAR regression (the code is: BS), as a baseline. TM_STA combines VAR (the code is: BS) with GARCH (the code is: EX) to evaluate the performance of handling unexplained fluctuation in high-frequency time series. Finally, we incorporate the social influence model (the code is: SI) into TM_STA to observe how social influence will affect the performance in a high-frequency time series.

In order to assess how accurately the proposed model can predict trends, the proposed TM_STA model was used to run a simulation of the trading processes on the Hushen-300 index (we assumed the value of the indicator was the share price and based our tactics on the analysis above. The more returns we can obtain from the simulation, the more satisfactory the proposed model is considered to be).

The trading system being evaluated has basic buy and sell functions. In addition, we can deploy different strategies in the system to give buy or sell signals. The strategies make decisions according to the December time series of the Hushen-300 index, which serves as the testing data. October and November data are used to train our model for strategy design. We evaluate according to the five trading strategies listed below:

- VAR strategy (BS): we only use the VAR model to make predictions.
- Simple Trading Tactics: when an increase in the Hushen-300 Index was detected, the simulation sold; when there was a decrease, it held (Simple Strategy).
- Simple Hedge Tactics: when changes in the Hushen-300 Index are detected, we assign a different number of lags to make a Hedge Tactics-based Trading (Random Strategy).
- TM_STA (BS + EX) without the Social Influence model: we do not apply the SI model to optimize the original Tencent Mood data, EX is the GARCH model.
- TM_STA (BS + EX+SI) with the Social Influence model: we apply the SI model to optimize the original Tencent Mood data, EX is the GARCH model.

TABLE 6. Statistical analysis between positive sentiment percentage and price volatility.

	Morning price volatility (lags: 9 min)				Afternoon price volatility (lags: 6 min)			
Positive Sentiment Percentage	Num	Avg Dev	Max Dev	Min Dev	Num	Avg Dev	Max Dev	Min Dev
>80%	84	**18**	**66**	−53	99	12	**46**	−38
>75%	535	11	63	−55	661	**16**	42	−42
>70%	346	−8	58	−45	228	7	44	**−35**
>65%	113	4	61	**−42**	132	−10	41	−37
>60%	41	−9	46	−45	63	−15	36	−41

DOI: 10.1002/asi

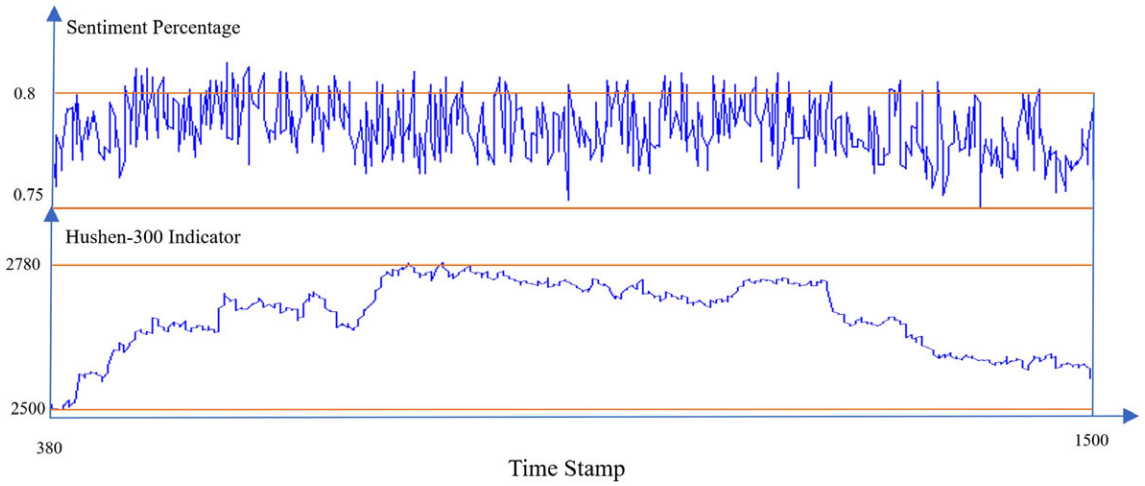

FIG. 6. The fluctuation analysis of Hushen-300 index during timestamp 380 and 1,500. [Color figure can be viewed at wileyonlinelibrary.com]

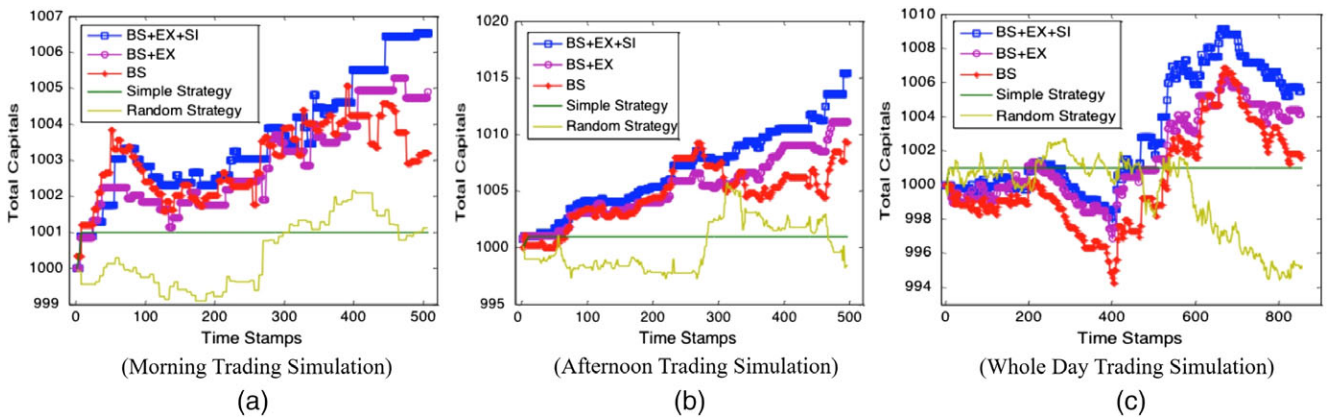

FIG. 7. Simulation results of different strategies. [Color figure can be viewed at wileyonlinelibrary.com]

Our proposed strategy is BS + EX+SI. The strategy is mainly based on the correlations detected in Table 5. For example, for morning trading the sentiment percentage at time t is $s(t)$, the Hushen-300 index at time t is $y(t)$, the predicted Hushen-300 at $t + 9$ is $\tilde{y}(t+9)$. Then if $\tilde{y}(t+9)-y(t) > threshold$, the system will adopt buying behaviors. The initial capital is 1,000 and the results of the simulation are summarized in Figure 7. The values at the last timepoint are the final returns (accumulated returns) of each tactic.

In Figure 7, the proposed model (BS + EX+SI) outperformed the other baseline significantly. TM_STA with social influence (BS + EX+SI) improves 35% in the morning, 20% in the afternoon, compared with TM_STA without social influence (BS + EX). TM_STA without social influence (BS + EX) improves 56% in the morning, 33% in the afternoon, compared with baseline VAR regression (BS).

Conclusion

Correlations between social media and stock markets have been widely studied, but the factors behind the correlations are seldom considered, especially in the case of high-frequency time series, which are of considerable importance in the era of big data. Classic financial theory cannot explain the fluctuations that frequently occur in stock markets. By contrast, behavioral financial theory provides a new, psychological perspective on the mechanisms of short-term rises and falls in the stock market by analyzing investors' irrational behaviors. Existing research mainly uses financial indicators to measure investors' irrational behaviors, which are often considered incomplete. In our research, we assume that users directly express their opinions in Tencent Weibo, and that influential opinions are diffused through the whole social network. This reflects irrational behaviors in the stock market, and Tencent users who use financial terms to express their moods are likely to be real investors. So we use a sentiment detection algorithm to detect Tencent moods to simulate investor sentiments, and we use social influence (SI) to model behavior diffusion in the stock market. Experimental results show that SI-based Tencent finance-related moods can better explain fluctuations in the stock market, which, according to behavioral financial theory, are caused by investors' irrational behaviors during a short decision time period

DOI: 10.1002/asi

(high-frequency time series). The findings are meaningful because we were able to use new data sources to support the existing theory and provide new perspectives with more explainable results to help understand stock market behavior. The research is not only of potential value to investors, but may also offer useful insights to financial supervision departments wishing to monitor and analyze risks at a micro-level.

However, several problems still need to be addressed: first, the detected correlations are considered to be weak connections from statistical economic angles, which means that unstable factors exist for the designed model; for example, different kinds of financial news may cause different social influences; in our current studies, we do not distinguish them. Second, the capability for addressing the issue of data sparsity is still limited. One possible area for improvement would be the incorporation of more data sources into the current model by applying a transfer learning approach; in addition, other efficient algorithms for content understanding and complex relations modeling, such as deep learning, might also be introduced to improve performance.

Acknowledgments

This work is supported by the Chinese National Key Foundation Research (61533018), Chinese National Youth Foundation Research (61702564), and Scientific Research Foundation (20000-18831102).

References

Aldridge, I., & Krawciw, S. (2017). Real-time risk: What investors should know about Fintech, high-frequency trading and flash crashes. Hoboken, NJ: Wiley ISBN: 978-1-119-31896-5.

Andersen, T., Bollerslev, T., & Cai, J. (2000). Intraday and interday volatility in the Japanese stock market. Journal of International Financial Markets Institutions and Money, 10, 107–130.

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

Bakshi, R.K., Kaur, N., Kaur, R., & Kaur, G. (2016). Opinion mining and sentiment analysis. In Proceedings of the third International Conference on Computing for Sustainable Global Development. New Delhi, India. 31 October.

Barberis, N., Huang, M., & Santos, T. (2002). ProsPect theory and asset prices. The Quarterly Journal of Economics, 116, 1–53.

Bauman, W.S. (1967). Scientific investment analysis: Science or fiction? Financial Analysis Journal, 1, 93–97.

Benthaus, J. & Beck, R. (2015). It's more about the content than the users! The influence of social broadcasting on stock markets. AIS Electronic Library, 25(45), 6056-6066. Retrieved from http://aisel.aisnet.org/ecis2015_cr/17/.

Blanchard, O.J., & Watson, M. (1982). Bubbles, rational expectations and financial markets. In P. Wachtel (Ed.), Crises in the economic and financial structure (pp. 295–316). Lexington, MA: D.C. Heathand.

Bollegala, D., Mu, T., & Goulermas, J.Y. (2016). Cross-domain sentiment classification using sentiment sensitive embeddings. IEEE Transactions on Knowledge & Data Engineering, 28(2), 398–410.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics, 31(3), 7–27.

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., & Weber, I. (2012). Web search queries can predict stock market volumes. PLoS One, 7, e40014.

Borzykowsi, B. (2011). Twitter-tracking hedge fund beats market. Canadian Business. Retrieved from http://www.canadianbusiness.com/blogs-and-comment/twitter-tracking-hedge-fund-beats-market/

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.

Brown, G.W. (1999). Volatility, sentiment and noise traders. Financial Analysts Journal, 55(2), 82–90.

Burrell, O.K. (1951). Possibility of an experimental approach to investment studies. The Journal of Finance, 6(2), 211–219.

Carol, C. (2012). How to keep markets safe in the ear of high-speed trading. Chicago Fed Letter, 303, p. 1. Retrieved from http://xueshu.baidu.com/s?wd=paperuri%3A%28d83598090980e026f2a71ca5b88b17f8%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http%3A%2F%2Fconnection.ebscohost.com%2Fc%2Farticles%2F80238692%2Fhow-keep-markets-safe-era-high-speed-trading&ie=utf-8&sc_us=913640021981926266.

Chen, F., Bai, X., & Zhan, S. (2014). A study on recursive neural network based sentiment classification of Sina Weibo. IEEE International Conference on Trust (pp. 681–685).

Chen, W., Cai, Y., & Lai, K. (2016). Weibo mood towards stock market. In International Conference on Database Systems for Advanced Applications (pp. 3–14).

Chib, C., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. The American Statistician, 49(4), 327–335.

Chib, S., Greenberg, E., & Chen, Y. (1998). MCMC methods for fitting and comparing multinomial response methods. Social Science Research Network. OLIN-97-15. SSRN Retrieved from https://ssrn.com/abstract=61445.

Corea, F., & Cervellati, E.M. (2015). The power of microblogging: How to use twitter for predicting the stock market. Eurasian Journal of Economics & Finance, 3(4), 1–7.

Da, Z., Engelberg, J., & Gap, P. (2015). The sum of all fears investor sentiment and asset prices. Review of Financial Studies, 28(1), 1–32.

DeMarzo, P.M., Vayanos, D., & Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. Quarterly Journal of Ecomonics, 118(3), 909–968.

Dennis, P., & Stewart, M. (2002). Risk-neutral skewness: Evidence from stock options. Journal of Financial & Quantitative Analysis, 37, 471–493.

Devi, L., Palaniappan, S., & Kumar, P. (2015). Tweet sentiment classification using an ensemble of machine learning supervised classifiers employing statistical feature selection methods. In Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (pp. 1–13).

Dong, Y., Tang, J., Wu, S., Tian, J., Chawla, N., Rao, J., & Cao, H. (2012). Link prediction and recommendation across heterogenous social networks. In Proceedings of 2012 I.E. International Conference on Data Mining (ICDM'12).

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50, 987–1007.

Harris, L. (1986). A transaction data study of weekly and Intradaily patterns in stock returns. Journal of Financial Economics, 16, 99–117.

Hirshleifer. D., Subrahmanyam, A., & Titman, S. (2002). Feedback and the success of irrational investors. In Ohio State University Working Paper.

Karabulut, Y. (2013). Can Facebook predict stock market activity?. AFA 2013 San Diego Meetings Paper. Retrieved from https://ssrn.com/abstract=2017099 or https://doi.org/10.2139/ssrn.2017099

Kwak, H., Lee, H.P., & Moon, S. (2010). What is Twitter, a social network or a news media? In Proceedings of the 19th International Conference on World Wide Web (pp. 591–600).

Lashgari, M. (2000). The role of TED spread and confidence index in explaining the behavior of stock prices. American Business Review, 18, 9–11.

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, 2277–2299.

Lee, C.M.C., Shleifer, A., & Thaler, R.H. (1991). Investor sentiment and the closed-end fund puzzle. The Journal of Finance, 46(1), 75–109.

Li, D., Shuai, X., Sun, G., Tang, J., Ding, Y., & Luo, Z. (2012). Mining topic-level opinion influence in microblog. In Proceedings of the Twenty-First Conference on Information and Knowledge Management (CIKM'12) (pp. 1562–1566).

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Mining Text Data (pp. 459–526). Berlin: Springer.

Liu, L., Tang, J., Han, J., Jiang, M., & Yang, S. (2010). Mining topic-level influence in heterogeneous networks. In Proceedings of the Nineteenth Conference on Information and Knowledge Management (CIKM'10) (pp. 199–208).

Long, J.B.D., Shleifer, A., Summers, L.H., & Waldmann, R.J. (1990). Positive feedback investment strategies and destabilizing rational speculation. The Journal of Finance., 45, 379–395.

Mao, Y., Wang, B., Wei, W., & Liu, B. (2012). Correlating S&P 500 stocks with Twitter data. In HotSocial'12, August, 12, Beijing, China.

Neal, R., & Wheatley, S.M. (1998). Do measures of investor sentiment predict returns? The Journal of Financial and Quantitative Analysis, 33, 523–547.

Nguyen, L., Wu, P., Chan, W., Wei, P.W., & Zhang, J. (2012). Predicting collective sentiment dynamics from time-series social media. In Proceedings of Workshop on Issues of Sentiment Discovery and Opinion Mining at The 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '12).

Nofer, M., & Hinz, O. (2015). Using twitter to predict the stock market: Where is the mood effect. Business & Information Systems Engineering, 57(4), 229–242.

Nofsinger, J.R. (2005). Social mood and financial economics. Journal of Behavioral Finance, 6(3), 144–160.

Pagolu, C., Challa, K.N.R., Panda, G., & Majhi, B. (2016). Sentiment analysis of twitter data for predicting stock market movements. In International Conference on Signal Processing, Communication, Power and Embedded System.

Patel, V., Prabhu, G., & Bhowmick, K. (2015). A survey of opinion mining and sentiment analysis. International Journal of Computer Applications, 131(1), 24–27.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems, 89, 14–46.

Scharfstein, D.S., & Stein, J.C. (1990). Herd behavior and investment, 80(3), 465–479.

Simkins, S.P. (1994). Do real business cycle models really exhibit business cycle behaviour? Journal of Monetary Economic, 33, 381–404.

Slovic, P. (1972). Psychological study of human judgment: Implications for investment decision making. TheJournal of Finance, 27(4), 779–801.

Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., & Li, P. (2011). User-level sentiment analysis incorporating social networks. In Proceedings of the Seventeenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11) (pp. 1397–1405).

Tang, D., Qin, B., & Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification. In Conference on Empirical Methods in Natural Language Processing (pp. 1422–1432).

Tang, J., Sun, J., Wang, C., & Yang, Z. (2009). Social influence analysis in large-scale networks. In Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009) (pp. 807–816).

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). Aminer: Extraction and mining of academic social networks. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 990–998). Las Vegas. 24–27. August.

Telock, P.C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139–1168.

Teruo, N. (2000). Bayesian analysis of ARMA-GARCH models: A Markov chain sampling approach. Journal of Econometrics, 95, 57–69.

Turney, P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (pp. 417–424). Philadelphia: Association for Computational Linguistics.

Tversky, A., & Kanneman, D. (1992). Advances in prospect theory. Cumulative representation of uncertainty. Journal of Risk and Uncertainty, 5, 297–323.

Wang, Z., Tong, V.J.C., & Chin, H.C. (2014). Enhancing machine learning methods for sentiment classification of web data. In the proceedings of 10th Asia Information Retrieval Societies Conference, AIRS 2014, Kuching, Malaysia, December 3-5.

Wood, R.A., McInish, T.H., & Ord, J. (1985). An investigation of transactions data for NYSE stocks. The Journal of Finance, 40, 723–739.

Wu, S., Fang, Z., & Tang, J. (2012). Accurate product name recognition from user generated content. (ICDM Contest) In Proceedings of ICDM 2012 Contest (pp. 874-877). Retrieved from http://www.gabormelli.com/Projects/CPROD1workshop/

Yang, C., & Zhou, L. (2015). Investor trading behavior, investor sentiment and asset prices. The North American Journal of Economics & Finance, 34, 42–62.

Yang, J., & Yecies, B. (2016). Mining Chinese social media UGC: a big-data framework for analyzing Douban movie reviews. Journal of Big Data, 3(1), 1–23.

Yu, H., & Hatzivassloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP'03) (pp. 129–136). Association for Computational Linguistics.

Zhang, X., Fuehres, H., & Gloor, P.A. (2011). Predicting stock market indicators through twitter "I hope it is not as bad as I fear". Procedia-Social and Behavioral Sciences, 26, 55–62.

Zhang, Y., Chen, M., Liu, L., & Wang, Y. (2017, June). An effective convolutional neural network model for Chinese sentiment analysis. In AIP Conference Proceedings, 1836(1).

Zhuang, H., Tang, J., Tang, W., Lou, T., Chin, A., & Wang, X. (2012). Actively learning to infer social ties. Data Mining and Knowledge Discovery, 25(2), 270–297.

Zou, H., Tang, X., Xie, B., & Liu, B, (2015). Sentiment classification using machine learning techniques with syntax features. In Proceedings of International Conference on Computational Science and Computational Intelligence (CSCI). Las Vegas, NV, USA. 7-9 Dec.