Received: 25 February 2020 | Revised: 10 August 2020 | Accepted: 10 August 2020
DOI: 10.1002/ijfe.2260

RESEARCH ARTICLE

WILEY

Can sentiments on macroeconomic news explain stock returns? Evidence form social network data

Yingying Xu[1] | **Jichang Zhao**[2]

[1]School of Economics and Management, University of Science and Technology Beijing, Beijing, China

[2]School of Economics and Management, Beihang University, Beijing, China

Correspondence
Yingying Xu, School of Economics and Management, University of Science and Technology Beijing, Beijing, China.
Email: yingxu21@ustb.edu.cn

Funding information
Fundamental Research Funds for the Central Universities, Grant/Award Number: 06500106yt5bg; National Natural Science Foundation of China, Grant/Award Number: 71873014; USTB-NTUT Joint Research Program, Grant/Award Number: 06310058

Abstract
Macroeconomic factors and sentiments affect investors' decisions and thus stock returns. However, do sentiments on macro-economic news explain stock returns? This article proposes a theoretical model to explain the relationship between stock returns and the market misperception which is driven by investors' sentiments. Then, microblogs regarding the macro-economy posted on Sina Weibo, a mainstream Chinese Social Network Site, are extracted to measure investors' macroeconomic sentiments (IMSs) through machine learning approaches. A preliminary event study suggests that IMSs capture the development of influential macroeconomic events. Empirical results demonstrate that orthogonalized IMSs including anger, disgust, fear, joy and sadness exert heterogeneously significant effects on the Shanghai Composite Index (SHCI), and no reverse effect is found. Thus, the IMS contains additional information related to the macro-economy; but cannot be explained by macroeconomic factors. IMSs improve the in- and out-of-sample predictabilities of SHCI returns. Thereby, investors' sentiment can be an important channel through which the macro-economy affects the stock market.

KEYWORDS
Macro-economy, predictability, sentiments, stock return, Weibo

1 | INTRODUCTION

Stock market prediction has attracted much attention theoretically, empirically and experimentally. A plethora of models and approaches have emerged to address this issue (e.g., Goetzmann & Jorion, 1993; Lewellen, 2004). Numerous financial and macroeconomic factors have been employed (see, inter alia, Fama & French, 1988; Nofer & Hinz, 2015; Hong, Chen, O'Brien, & Ryan, 2018). Substantial evidence supports the predictability of stock returns, particularly the effects of macroeconomic factors, whereas minimal literature considers the channel through which the macro-economy affects stock markets, which is of significance for investment strategies and policy managements.

News most certainly influence stock returns, investors' sentiments also play an equally important role (Bollen, Mao, & Zeng, 2011). As discussed by behavioral finance and neuro-finance researchers, sentiments, in addition to information, play a significant role in human decision-making process (Dolan, 2002; Nofer & Hinz, 2015). Empirical evidence on stock return predictability is mixed when public sentiments are considered (Xu, Liu, Zhao, & Su, 2017). By contrast, many studies suggest that macroeconomic factors work in forecasting stock market performances (Neely, Rapach, Tu, & Zhou, 2014). It is therefore reasonable to assume that the macro-economy can drive stock market movements through investors' sentiments. In other words, sentiments induced by changes in the macro-economy, that is, investors' macroeconomic sentiment (IMS), rather than public emotions caused by daily life events such as private life, are more likely to affect the stock market.

Sentiment tracking techniques that extract indicators of agents' mood directly from Social Network Sites (SNS) such as Twitter, Facebook and Sina Weibo have made significant progress over the past decade (Bollen et al., 2011; Fan, Zhao, Chen, & Xu, 2014). Millions of blogs are posted on SNSs every day, which provides an accurate representation of certain groups such as investors. Machine learning techniques and sentiment measurement techniques enable us to screen contents regarding the macro-economy and measure corresponding sentiments, thus providing reliable and scalable assessments of the IMS. On such basis, we attempt to research whether orthogonalized IMSs which mitigate effects of macroeconomic conditions affect stock markets and help predict stock movements. This will reveal whether the macro-economy can affect stock markets through investors' sentiments. Frequent policy changes and interventions introduce abundant non-market influences to Chinese stock markets which contain a significant quantity of individual investors. Thereby, the Chinese stock market provides an appropriate environment to investigate the relationship between IMSs and stock returns.

De Long, Shleifer, Summers, and Waldmann (1990) provides a concrete model which shows that asset prices may deviate from fundamental values because of irrational waves of sentiments. We promote a theoretical model according to (Christou, Cunado, Gupta, & Hassapis, 2017) to explain how IMSs affect stock returns and further empirically investigate such effects. Given significant heterogeneity among investors' sentiments, it is of particular interest to test whether effects of different sentiments are homogeneous on stock returns. Additionally, we estimate whether the predictability of stock returns can be improved using IMSs. To achieve these goals, we collect approximately 680,123 of blogs regarding the macro-economy with explicit emotions from April 1, 2013 to October 31, 2014, before Chinese stock market entered the turmoil in 2015. We then classify blogs into five sentiments, that is, anger, disgust, fear, joy, and sadness. Before estimating the effects of IMSs on stock returns, we illustrate that sentiments can tract macroeconomic changes. IMSs are not simple reflections of the macro-economy because they actually convey perceptions of agents after processing information related to the macro-economy. Therefore, we measure the orthogonalized sentiments that mitigate effects of macroeconomic conditions, and then adopt regression analysis which controls some macroeconomic variables to measure the net effects of investors' IMSs on the stock return of Shanghai Composite Index (SHCI). Pervasive significant long-term correlations between stock returns and IMSs are demonstrated, except the disgust sentiment. The results based on the first principal component of five orthogonalized sentiments and results using holiday data confirm the robustness of our findings. Finally, we compare the out-of-sample predictability of IMSs through linear and nonlinear models and show that IMSs, particularly joy and sadness, contain useful information in forecasting the SHCI return. Overall, the empirical results agree with our theoretical model and provide evidence supporting that investors' sentiments based on macro-economic news explain stock markets, and such sentiments show leading effects on stock returns. Information, specifically sentiments caused by changes in the macro-economy such as monetary policies, on SNSs which can be easily collected by investors can be used to predict the movements of stock returns.

Our contributions to the literature are threefold. First, this article proposes a model describing the relationship between stock returns and the market misperception, which is mainly driven by investors' sentiments based on the noise trader model of De Long et al. (1990). In our model, the market misperception which can be caused by noise traders' judgment about (true or pseudo) information of the macro-economy determines the stock price and return. The theoretical model describes that noise traders' perceptions, that is, sentiments, affect their investment decisions. Meanwhile, under interaction of the number of noise traders, the market misperception, that is, the average misperception of all noise traders, determines the equilibrium price and return of sophisticated investors and noise traders. Given that individual psychological factors and rumors are more likely to impact certain investors' misperceptions or specific stocks, their effects on the entire stock market offsets each other. Thereby, we further infer that misperceptions caused by changes in macroeconomic conditions, which exert extensive effects on investors and stocks, can drive stock returns.

Second, the macroeconomic sentiment is computed daily through machine learning of microblogs posted by millions of users of Sina Weibo. Survey-based measures of sentiments are criticized because respondents may belie their real thoughts. By contrast, the IMS is based on what users "actually" think or feel (Karabulut, 2013). Furthermore, survey-based measures are generally available at monthly or quarterly frequencies and are expensive and time-consuming to conduct. Most recent studies using online information to measure sentiments focus on public sentiments of all users, thus may lead to significant bias to investors' sentiments. Through screening microblogs regarding the macro-economy, we narrow down the users to the group more closely related to the stock market, thus bring us closer to the source of investors' sentiments rather than public sentiments.

Third, in contrast to previous studies that research polarized sentiments, that is, positive or negative, we consider detailed sentiments including anger, disgust, fear, joy and sadness. Meanwhile, to mitigate possible correlations between sentiments and macro-economic factors, we use the orthogonalized detailed sentiments and the first principal component of them to test the net effects of IMSs on stock returns. Accordingly, we can investigate whether the IMS contains additional information that the macroeconomic factors cannot provide. Particularly, we demonstrate that different sentiments exert heterogeneous effects on the stock market and show various predictabilities in forecasting stock returns.

2 | LITERATURE REVIEW

This article is mainly related to two strands of literature: how sentiments affect asset markets and the effects of macroeconomic factors on stock markets. The theoretical literature regarding how sentiments affect asset price can be traced back to the seminal work of De Long et al. (1990), who introduce the concept of "irrational noise traders" in the modeling of asset pricing. In their model, two types of investors exist, that is, rational arbitrageurs and noise traders who are not fully rational. The equilibrium price of financial markets depends on rational expectations of rational arbitrageurs and irrational expectations of noise traders who are subject to exogenous sentiments that are overly optimistic or pessimistic. Consequently, the asset price may deviate from fundamental values under waves of noise traders' sentiments, thus showing the effect of sentiments on stock prices. Following the pioneering work of De Long et al. (1990), many studies attempt to analyze and measure the effects of investors' sentiments on asset markets (e.g., Garcia, 2013).

Traditionally, sentiments can be measured directly based on survey or indirectly based on market. For example, the trading amount, market volatility, the closed-end fund discounts, net mutual fund redemption, the ratio of odd-lot sales to purchases, value-weighted dividend premiums, the number of Initial Public Offerings (IPOs), the average first-day IPO returns, the equity share in new issues and turnover are indirect measurements of sentiments (Baker & Wurgler, 2006). Large surveys of sentiments over representative samples of the population are generally expensive and time-consuming to conduct, and the accuracy of indirect measurement is limited given that the chosen indicators are expected to be correlated with sentiments (Bollen et al., 2011).

Recently, sentiment-tracking techniques that extract public sentiments directly from social media content have made significant progress. Sentiments have been tentatively explored in relationships with economic indicators. Sentiment indicators based on online big data have been demonstrated to be useful in capturing investors' sentiments. For example, Nofer and Hinz (2015) analyze the relationship between simple and follower-weighted sentiments extracted from Germany Twitter and stock markets and find that follower-weighted sentiments are positively related to stock index returns. Oh and Sheng (2011) measure sentiments through 72,221 microblog postings for 1,909 stock tickers and demonstrate their predictive power for simple and market adjusted returns. They show that the predictive accuracy is consistent with the under-reaction hypothesis observed in behavioral finance, thus providing support for the model of irrational investor sentiment. Vu, Chang, Ha, and Collier (2012) identify positive and negative sentiments of tweets posted on Twitter and document sentiments' predictive power in forecasting stock prices. Textual information on other social media is also used to measure public sentiments, for example, Gilbert and Karahalios (2010) estimate anxiety from blogs posted on Live Journal and confirm that such index predicts downward pressure on the Standard & Poor's 500 index. Karabulut (2013) and Siganos, Vagenas-Nanos, and Verwijmeren (2014) adopt a Gross National Happiness (GNH) index which measures the difference between positive and negative sentiments extracted from blogs on Facebook to predict daily stock returns and trading volumes. Mao, Counts, and Bollen (2015) compare the forecasting performance of traditional surveys (Investor Intelligence and Daily Sentiment Index) against online sentiment indicators (Twitter Investor Sentiment, Negative News Sentiment, Tweet & Google Search volumes of financial terms, Twitter Investor Sentiment, Negative News Sentiment and Tweet & Google Search volumes of financial terms). They conclude that survey sentiment indicators appear to be insignificant in predicting financial market values. By contrast, certain online sentiment indicators such as Google Insight Search volumes on financial search queries, Twitter Investor Sentiment, and the frequency of occurrence of financial terms on Twitter have predictive values.

Studies about Chinese market support close relationships between sentiments extracted from social media and stock markets. For instance, Cheng and Lin (2013) show short-term effects of Bullishness and Bearishness indices on stock market returns, but have no prediction power. Zhou, Zhao, and Xu (2016) note that the stock market in China can be competently predicted by emotions extracted from Sina Weibo, for example, disgust, fear, joy and sadness.

The above studies recommend a complimentary investing approach using user-generated content and validate the effects of sentiments extracted from SNSs on asset markets. Nevertheless, most research categorize blogs into positive and negative sentiments, thus ignoring detailed dimensions. Negative emotions such as anger, disgust and sadness are more applicable in real world scenarios, that is, emergency tracking and abnormal event detection (Zhao, Dong, Wu, & Xu, 2012). The research using Twitter finds that public sentiments extracted from massive information, particularly calm and the combination of calm and happiness can predict the closing price of DJIA (Bollen et al., 2011), whereas sentiment polarities cannot provide useful predictions of the Hong Kong Stock Exchange prices (Li, Xie, Chen, Wang, & Deng, 2014). Other studies, for example, Dong, Chen, Qian, and Zhou (2015), Zhou et al. (2016), and Xu et al. (2017) provide further evidence supporting the predictive power of detailed sentiments in forecasting stock markets.

The other strand of literature relating to this article concern the effects of macroeconomic factors on stock markets, particularly on the predictability (e.g., Hong et al., 2018; Wang, Wei, Wu, & Yin, 2018). However, conclusions maintain controversial. Various macroeconomic factors have been investigated. For example, short-term interest rate, industrial output, inflation, unemployment rate, oil price and so forth. Recent studies construct new indicators relating to news to analyze the effect of macroeconomic factors on stock returns. Christou et al. (2017) estimate the effect of economic policy uncertainty (EPU) on stock market returns in PacificRim countries and illustrate a negative impact of the EPU which is based on media disclosure frequency on stock returns. Hammerschmid and Lohre (2018) extract three principle components from 14 macroeconomic factors and demonstrate that they improve the in-sample predictive power in forecasting stock returns when combined with a technical index. However, they also point out that the out-of-sample predictive effect is time-varying. Carabias (2018) suggests that macroeconomic news help predict the end-of-quarter realized earnings of companies.

The mechanism that macroeconomic factors and news affect stock market can be explained from the perspective of investors' risk decision process. As noted by Liu (2017), many macroeconomic factors, for example, economic growth, inflation, interest rate, money supply and trade policy explain investors risk return trade-off. Consequently, the influence of macro-economy on investors' sentiments is likely to be transmitted to agents' risk attitude, which further affects their decision process. However, to the best of our knowledge, no literature covers this topic.

3 | OVERLAPPING GENERATIONS MODEL

De Long et al. (1990) introduce the concept of “irrational noise trader” in asset pricing models and demonstrate that noise traders' sentiment affects stock prices. Following De Long et al. (1990), the behavioral finance provides further proof that financial decisions are driven by sentiments (Al-Hajieh, Redhead, & Rodgers, 2011). Align with De Long et al. (1990), we assume that two kinds of traders exist in the market, that is, noise traders (denoted n) and sophisticated investors (denoted s). Noise traders believe that they have special information about future prices of risky assets and accordingly formulate their investment strategies. Such incorrect beliefs can be based on signals from technical analysts, stockbrokers, economic consultants, or rumors in media reports and social networks. Sophisticated traders exploit noise traders' irrational misperceptions and adopt contrarian investment strategies. Specifically, sophisticated traders sell when noise traders push prices up and buy when noise traders depress prices. The gambling between noise traders and sophisticated investors push asset prices toward fundamentals; but not all the time.

The resources agents invest are exogenous since there is no first-period consumption, no labor supply decision, and no bequest. Agents make their investment decisions on two assets, that is, the risk-free asset (denoted F) in perfectly elastic supply and pays a fixed dividend r and the risk asset (denoted R) which pays an uncertain dividend $r + \epsilon_t$. Thus, the fixed real dividend r is the riskless rate, and ϵ_t is serially independent, normally distributed with zero mean and constant variance σ_ϵ^2. For simplicity, we further assume that σ_ϵ^2 is uncorrelated with noise traders' opinions ρ_t. The risk asset is not in elastic supply, but is in fixed and unchangeable quantity, normalized at one unit. Usually, the risk asset can be interpreted as aggregate equities and noise trader risk is market-wide rather than idiosyncratic.

We assume that noise traders present in the model in measure μ_t and sophisticated investors present in measure $1 - \mu_t$. Similar with De Long et al. (1990), we further assume that the percentage of noise traders is an identically distributed normal random variable, that is, $\mu_t \sim N(\mu^*, {\sigma_\mu}^2)$. When agents are young, they choose their portfolios by maximizing individual perceived expected utility which depends on personal expectations of the mean of the distribution of the price of R, that is, p_{t+1}. A sophisticated investor holds a rational expectation and determines individual portfolio given that distribution. By contrast, a noise trader possesses a misperception of the expected price of R, *that is*, ρ_t and $\rho_t \sim N(\rho^*, \sigma_\rho{}^2)$. The misperception of the noise trader can be disturbed by pseudo signals derived from various aspects such as over or insufficient interpretation of changes in macroeconomic factors, rumors and individual psychological factors. However, compared with changes in the macro-economy, which can be misinterpreted by media reports to a certain direction (Dräger, 2015), rumors and psychological factors are more likely to impact certain stocks or investors rather than result in pervasive influences.

Furthermore, the percentage of noise traders is correlated with signals, particularly the macro-economy. Different from De Long et al. (1990), we assume that the percentage of noise traders is related to signals for two reasons. First, macroeconomic fluctuations may lead noise traders enter or leave the stock market according to their risk appetites. Second, the past relative performance of excess return or utilities is closely related to the macro-economy. For example, the moderately loose monetary policy and expansionary fiscal policy in 2015 facilitated the inflow of money into the stock market, thus rose the price. Meanwhile, China's economy stepped into the "new normal" phase with a lower GDP growth. The divergence between economic fundamentals and stock prices, accelerated by media's improper guides, prompted the stock disaster in June 2015. Then, lots of retail investors fled from the stock market to the real estate market, which was boosted by the destocking policy at that time. Consequently, the percentage of noise traders is likely to be correlated with the macro-economy. Therefore, we further assume that the average noise trades' misperception, that is, the market misperception $q_t = \rho_t \mu_t \sim N(q^*, \sigma_q{}^2)$. The mean market misperception q^* is a measure of the average excess expectation of all noise traders to the rational expectation of sophisticated investors, and $\sigma_q{}^2$ is the variance of noise traders' misperceptions of the expected return per unit of the risky asset. Noise traders thus maximize their own expectation of utility given the next-period dividend, the one-period variance of p_{t+1} and their false belief that the distribution of the price of R next period has mean ρ_t above its true value.

$$U = -e^{-(2\gamma)w}, \tag{1}$$

where γ is the coefficient of absolute risk aversion. Given that the return of holding a unit of the risky asset is assumed to be normally distributed, maximizing the expected value of Equation (1) is equivalent to maximizing

$$\bar{w} - \gamma \sigma_w{}^2, \tag{2}$$

where $\bar{w}$ is the expected final wealth, and $\sigma_w{}^2$ is the one-period-ahead variance of wealth. Therefore, a representative sophisticated investor chooses the amount λ_t of the risky asset R to maximize his/her utility

$$E(U) = \overline{w} - \gamma \sigma_w{}^2 = c_0 + \lambda_t^i \left[r + E_t p_{t+1} - p_t(1+r) - \gamma \left(\lambda_t^i\right)^2 E_t\left(\sigma^2_{p_{t+1}}\right)\right], \tag{3}$$

where c_0 is a function of first-period labor income, and $E_t p_{t+1}$ denotes that the expectation of p_{t+1} is taken at time t. We define $E_t\left(\sigma^2_{p_{t+1}}\right)$, the one-period variance of p_{t+1}, as follows:

$$E_t\left(\sigma^2_{p_{t+1}}\right) = E_t\left\{(p_{t+1} - E_t p_{t+1})^2\right\}, \tag{4}$$

The representative noise trader maximizes

$$\begin{aligned} E(U) = \overline{w} - \gamma \sigma_w{}^2 = c_0 \\ + \lambda_t^n \left[r + E_t p_{t+1} - p_t(1+r) - \gamma \left(\lambda_t^n\right)^2 E_t\left(\sigma^2_{p_{t+1}}\right)\right] \\ + \lambda_t^n (\rho_t), \end{aligned} \tag{5}$$

In Equation (5), $\lambda_t^n(\rho_t)$ captures the noise traders' misperception of the expected return from holding λ_t^n units of the risky asset. All young agents determine their portfolios by maximizing Equation (3) and Equation (5). We thus get the following results of λ_t^i and λ_t^n which represent the quantities of the risky asset purchased by sophisticated investors and nosier traders, respectively.

$$\lambda_t^i = \frac{r + E_t p_{t+1} - (1+r)p_t}{2\gamma\left[E_t\left(\sigma^2_{p_{t+1}}\right) + E_t\left(\sigma^2_\epsilon\right)\right]}, \tag{6}$$

$$\lambda_t^n = \frac{r + E_t p_{t+1} - (1+r)p_t + \rho_t}{2\gamma\left[E_t\left(\sigma^2_{p_{t+1}}\right) + E_t\left(\sigma^2_\epsilon\right)\right]}, \tag{7}$$

The two-generations model assumes that old agents sell their holdings of R for price p_{t+1} to the new young, and so the demands of the young must sum to one in equilibrium. Equations (6) and (7) thus imply that

$$p_t = \frac{1}{1+r}\left\{ r + E_t p_{t+1} - 2\gamma\left[E_t\left(\sigma^2_{p_{t+1}}\right) + E_t\left(\sigma^2_\epsilon\right)\right] + \mu_t \rho_t \right\}, \tag{8}$$

We further impose the requirement that the unconditional distribution of p_{t+1} be identical to the distribution of p_t and obtain the following expression of p_t:

$$p_t = 1 + \frac{q_t - q^*}{1+r} + \frac{q^*}{r} - \frac{2\gamma}{r}\left[E_t\left(\sigma^2_{p_{t+1}}\right) + E_t\left(\sigma^2_{\epsilon}\right)\right], \tag{9}$$

where the one-step-ahead variance of p_t is an unchanging function of the constant variance of a generation of market misperception q_t:

$$E_t\left(\sigma^2_{t_{t+1}}\right) = \frac{{\sigma_q}^2}{(1+r)^2}, \tag{10}$$

Finally, the price and return of R depend on public information about present and future market misperceptions and certain exogenous parameters $(r, \gamma, q^*, {\sigma_q}^2, \sigma^2_{\epsilon})$

$$p_t = 1 + \frac{q_t - q^*}{1+r} + \frac{q^*}{r} - \frac{2\gamma}{r}\left[\sigma^2_{\epsilon} + \frac{{\sigma_q}^2}{(1+r)^2}\right], \tag{11}$$

$$R_t = Ln(p_t) - Ln(p_{t-1}) = Ln\left(\frac{q_t + a}{q_{t-1} + a}\right) \propto \frac{q_t}{q_{t-1}}, a = 1 + r + \frac{1 + r - 2\gamma\left[{\sigma_q}^2 + (1+r)\sigma^2_{\epsilon}\right]}{r(1+r)}, \tag{12}$$

According to Equations (11) and (12), when the number of noise traders and individual misperceptions increase simultaneously, the equilibrium price rises and thus leads to higher return. Consequently, macroeconomic factors can affect the stock market through influencing the number of noise traders and their misperceptions. When noise traders are more optimistic, that is, perceiving a positive excess return of asset R, they tend to bid up the price and return. When they are more pessimistic, that is, perceiving a negative excess return of asset R, they tend to bid down the price and return. This overlapping generations structure can be used to model the effects of macroeconomic factors and the induced market misperceptions on stock prices and returns.

As defined in the overlapping generations model, noise traders' misperceptions are based on their pseudo-signals that are closely related to social interactions. They irrationally believe that these signals carry information and exhibit the fallacy of excessive subjective certainty in formulating investment strategies. Therefore, the market misperception can be induced by macroeconomic factors but is not necessarily consistent with fundamentals. This is what Keynes (1936) referred to as "animal spirits," a voluntary impulse rather than an idle impulse that drives agents' decisions. Both the number of noise traders and their misperceptions affect stock markets. Therefore, idiosyncratic sentiments related to private life are not likely to affect the market widely. Empirically, most indicators used in the literature consider investors' financial sentiments rather than private sentiments. For instance, survey-based indexes such as consumer confidence (Fisher & Statman, 2003) and financial market indicators such as Market Volatility Index (VIX, Karabulut, 2013). Sentiments generated from various aspects of the macro-economy, which have been demonstrated to exert significant effects on stock markets, are more likely to change the market misperception and thereby the stock market. For example, fiscal and monetary policies, unemployment, price indexes and so forth.

4 | DATA AND PRELIMINARY ANALYSIS

The messages posted on Sina Weibo and other SNSs represent investor's expectations to a large extent, which drive individual investment decisions. Social interactions of internet users are unavoidable, thus leading to mood contagion on SNSs (Guillory et al., 2011). Consequently, the spread of sentiments among users will diffuse expectations to other investors, thus impacting the stock market. One can therefore expect a close relationship between sentiments extracted from microblogs and stock market indicators. Meanwhile, the stock market performance may induce fluctuations of investors' sentiments and strengthen the correlation between them. The existence of such correlation is the foundation of our research. To investigate the effect of IMSs on stock markets, we promote it from event studies to a general research of time series.

4.1 | Sentiment classification

The collection and classification of macroeconomic relevant microblogs are proceeded in three steps. First, we use keywords searching through all accounts of Sina Weibo to collect macroeconomic related microblogs. The keywords include Gross Domestic Product (GDP), Gross National Product (GNP), Consumer Price Index (CPI), fiscal policy, monetary policy, tax, employment, interest rate, exchange rate, open market operation, central bank bills, deposit reserve ratio, rediscount rate, money supply and so forth. The information of microblogs is provided by Sina Weibo through its open Application Programming Interfaces (APIs). In the first step, we collect approximately 750,000 of microblogs regarding the macro-economy from April 2013 to October 2014 as the initial set. During this period, approximately 1,300 microblogs regarding various macroeconomic factors are posted every day.

Second, following Fan et al. (2014) and Xu et al. (2017), we adopt the incremental learning Naive Bayes classifier to classify sentiments of microblogs into five detailed categories: angry, disgust, fear, joy and sadness. Similar with tweets, the content of microblogs posted on Sina Weibo is limited to 140 characteristics. Many different approaches have been developed to measure sentiments of texts in recent years, for example, lexicon-based method (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011) and machine learning based solutions (Zhao et al., 2012). Zhao et al. (2012) train a fast Naive Bayes classifier on data from Sina Weibo for emotion classification and programmed a system of MoodLens for temporal and spatial sentiment patter discovery. The system is available from http://gana.nlsde.buaa.edu.cn/hourly_happy/moodlens.html.

Finally, we restructure the data collected in the first step and keep 680,123 microblogs in the final research sample. About 69,877 macroeconomic relevant microblogs contain no explicit emotion, thus are removed from our initial set. In the final set, there are 120 to 6,050 daily postings regarding macroeconomic conditions and each contains an explicit emotion, which means an average of 1,183 posts each day with a standard deviation of 588 messages. We further distinguish posts in trading days from those posted in holiday days (including weekends) in the following empirical studies. During trading days, there are 494,841 macroeconomic related microblogs with sentiments.

We research the stock return of SHCI which is measured by the log difference of closing prices. The sample period expires at the end of October 2014, when the soaring stock prices of financial sector led to the rapid rise of the market because of the effect of high leverage. The supervision measures of the China Securities Regulatory Commission (CSRC) drove the stock market to a period of violent fluctuations, thus making Chinese stock market ahead of fundamentals. The divergence between stocks and the economy triggered new extremes. The data are available in the Bloomberg database.

The ability of sentiments in capturing investors' moods is tested through applying microblogs relating to three important economic events in a time period corresponding to the event time frame. Three economic events are chosen for their significant effects, that is, the “fat finger” event of China Everbright Securities (CES, August 2013 to March 2015), the Made in China concept (November 2014 to March 2015) and the Fourth-Generation communication technology license (4G license, April 2014 to February 2015). Significantly different sentiments generate for the same event, thus indicating that the IMS is not a simple reflection of changes in the macro-economy. The IMS is unsurprisingly related to the macro-economy since it is assumed to be a channel through which the economy affects stock markets. However, agents are heterogeneous, thus making their interpretations of the same information being entirely different. The analyzes are not included in this paper for brevity, but are available upon request.

4.2 | Data

We summarize the statistical analysis of SHCI return and sentiments of six dimensions (the log values of five detailed sentiments and the total number) in Table 1. We also use the Z-score of the numbers of microblogs align with Bollen et al. (2011) in all the following empirical analysis. The results are similar with the results using log values. Therefore, results using log values are reported in the paper for brevity. The skewness of the SHCI return is positive, meaning that the stock return skews to the right. Meanwhile, the Kurtosis of the SHCI return is larger than that of normal distribution, indicating the fat-tail property of Chinese stock market. If investors are overconfidence and regret aversion, they tend to rely on their own information and under-react to new information, thus resulting to comparatively smaller changes in stock prices (Bakar & Yi, 2016). This will lead to leptokurtosis. However, when a market trend is established, or strong information emerges, investors over-respond to previously ignored information and generate greater stock fluctuations compared with when all investors are rational. Finally, the overconfidence and regret aversion of investors generate leptokurtosis and fat-tail in stock returns, which is exactly the properties shown in Chinese stock market. Therefore, the assumption that noise traders who are irrational exist in the stock market is consistent with characteristics of real markets.

Accordingly, joy appears to be the top sentiment, followed by anger, fear, disgust and sadness. The result implies that although negative sentiments dominate positive ones, the optimistic sentiment spreads throughout the market given the minimum, maximum and mean values of joy higher than other detailed sentiments. However, the volatility of joy is the largest, thus showing that changes in macroeconomic factors may lead to drastic fluctuations in optimistic sentiments. Similar to the SHCI return, IMSs skew to the right, that is, more data fell on the right of the mean value. In other words, investors' attention to the macro-economy which is represented by posted microblogs with sentiments can be particularly high sometimes.

TABLE 1 Statistical analysis of investors' sentiments and SHCI stock return

	Anger	Fear	Disgust	Joy	Sadness	Total	SHCI
Mean	5.499	5.102	5.001	6.032	4.509	7.016	0.020
Max.	6.727	7.1934	6.471	8.148	6.460	8.708	3.333
Min.	2.565	0.000	1.609	3.583	0.693	4.205	−5.445
SD	0.482	0.584	0.524	0.616	0.553	0.488	0.991
Skewness	−1.240	−1.564	−1.2691	−0.034	−0.759	−1.010	−0.349
Kurtosis	8.110	15.851	9.517	4.230	10.544	7.961	5.762
J-B	637.156	3,454.872	966.123	29.986	1,169.457	566.657	129.902
Probability	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: Max. and Min. denote the maximum and minimum of corresponding sentiments (the log values of numbers) and the stock return, respectively. J-B refers to the value of Jarque-Bera test.

TABLE 2 Correlations between IMS and macroeconomic factors

Panel A: Correlations with IMS						
	Anger	**Disgust**	**Fear**	**Joy**	**Sadness**	**IMS-P**
Anger	1.000					0.856[a]
Disgust	0.659[a]	1.000				0.800[a]
Fear	0.682[a]	0.621[a]	1.000			0.768[a]
Joy	0.537[a]	0.655[a]	0.528[a]	1.000		0.795[a]
Sadness	0.755[a]	0.629[a]	0.630[a]	0.697[a]	1.000	0.855[a]
e_t	0.019	−0.197[a]	−0.027	−0.071	0.113[b]	−0.14[a]
r_t	0.202[a]	−0.007	0.206[a]	0.126[a]	0.206[a]	0.190[a]
Panel B: Correlations with orthogonalized IMS						
	Anger$^{\perp}$	**Disgust$^{\perp}$**	**Fear$^{\perp}$**	**Joy$^{\perp}$**	**Sadness$^{\perp}$**	**IMS − P$^{\perp}$**
Anger$^{\perp}$	1.000					0.891[a]
Disgust$^{\perp}$	0.720[a]	1.000				0.866[a]
Fear$^{\perp}$	0.669[a]	0.632[a]	1.000			0.781[a]
Joy$^{\perp}$	0.673[a]	0.672[a]	0.543[a]	1.000		0.859[a]
Sadness$^{\perp}$	0.733[a]	0.668[a]	0.528[a]	0.770[a]	1.000	0.870[a]
e_t	−0.078	−0.044	−0.043	−0.055	−0.060	−0.024
r_t	0.019	−0.009	0.013	0.012	−0.021	0.013

Note: Panel A reports the Pearson correlations with five sentiments. Panels B summarizes the Pearson correlations with orthogonalized sentiments. The superscript ⊥ denotes orthogonalization. We regress each of the five sentiments on the current and first lagged exchange rates and interest rates with the first order autocorrelation term. The orthogonalized sentiments are the residuals from these regressions. IMS-P is the first principal component of the five orthogonalized sentiments. r_t and e_t represent SHIBOR and exchange rate, respectively.

[a]Statistical significance at the 1% level.

[b]Statistical significance at the 5% level.

4.3 | IMS and macroeconomic factors

Given the construction of the IMS, one expects that the sentiment should be correlated with various macroeconomic variables. Considering that sentiment is not the only channel through which macro factors influence the stock market, as acknowledged in the literature review, it follows that these alternative effects need to be controlled for if the sentiment channel is to be properly demonstrated. Thus, we control two important daily available macro variables, that is, the Shanghai interbank rate (SHIBOR) and the exchange rate of Chinese RMB quoted in the United States Dollars. Through controlling for these more direct effects, we can demonstrate whether

TABLE 3 Effects of trading days' IMSs on the SHCI return

	Dependent variable: SHCI return (R_t)							
R_{t-1}	0.097[a]	0.096[b]	0.095[a]	0.099[a]	0.095[a]	0.095[a]	0.097[a]	0.096[b]
$Anger^{\perp}$		0.323[b](5)						
$Disgust^{\perp}$			0.196(5)					
$Fear^{\perp}$				0.317[a](5)				
$Joy^{\perp}$					0.331[c](5)			
$Sadness^{\perp}$						0.278[c](5)		
$Total^{\perp}$							0.388[c](5)	
$IMS-P^{\perp}$								0.070[c](5)
r_t	−0.135[b]	−0.144[b]	−0.137[b]	−0.142[b]	−0.142[b]	−0.139[b]	−0.142[b]	−0.142[c]
e_t	−0.396	0.332	0.255	0.268	0.404	0.311	0.534	0.457
MON_t	0.137	0.124	0.141	0.151	0.130	0.144	0.132	0.134
FRI_t	0.137	0.147	0.149	0.164	0.151	0.144	0.157	0.156
JAN_t	−0.196	−0.173	−0.137	−0.155	−0.140	−0.165	−0.146	−0.142
Cons	2.840	−1.615	−1.164	−1.236	−2.061	−1.502	−2.869	−2.391
Adj. R^2	0.024	0.036	0.028	0.036	0.043	0.036	0.041	0.040

Note: This table summarizes the results of Equation (14) and adjust the residual error with HAC to eliminate the effects of heteroscedasticity. No ARCH effect exists in residual errors. In this table, IMSs are orthogonalized sentiments. r_t and e_t represent SHIBOR and exchange rate, respectively. The lags of IMS are in parentheses of the coefficients. The first column represents the regression results of Equation (14) without any IMS, but with control variables, calendar effects of January effect (JAN_t), Monday effect (MON_t) and Friday effect (FRI_t). $IMS-P^{\perp}$ denotes the first principal component of five orthogonalized sentiments. All variables are stationary without unit root, and all models are free of serial correlation problems.

[a]Significance at the 5% level.

[b]Significance at the 10% level.

[c]Significance at the 1% level.

the relationships are in fact due to sentiments or because of the correlations of IMSs to macro-economic variables.

Table 2 suggests that five sentiments are significantly positively correlated with each other, indicating that some common factors stimulate investors' emotion expressions. Furthermore, we find significant correlations between certain sentiments and macro-economic variables. Specifically, disgust and sadness are correlated with the exchange rate and all sentiments except disgust are related with the interest rate. The first principal component of five sentiments also shows significant correlations with these two macroeconomic factors. Thus, the orthogonalized IMSs are further calculated and used in the following empirical studies, which means that the sentiment variables are residuals of corresponding IMS mitigating the effects of SHIBOR and exchange rates. The Panel B in Table 2 reports correlations between orthogonalized sentiments and macroeconomic factors. As expected, the orthogonalized sentiments show no significant correlations with exchange rate and SHIBOR. Orthogonalized sentiments maintain significant correlations with each other, indicating that, apart from macroeconomic conditions, other common factors such as psychological and behavior paradigms are important drivers of investors' emotional changes. The first principal component of five orthogonalized sentiments displays no correlation with any macroeconomic factor. Aside from testing specific sentiments' effect on stock returns, we also investigate the overall impact of IMSs using the first principal component for robustness.

5 | EMPIRICAL RESULTS

5.1 | IMSs' effect on stock returns

To quantitatively determine the relations between stock returns and IMSs, we research the relationship between six dimensions of IMSs and the SHCI return using different methods. Firstly, we follow the seminal work of Nofer and Hinz (2015) to build an empirical model to test the effects of IMSs on stock returns. Most studies report a relationship between sentiments and stock market reactions on the next trading day (Karabulut, 2013). However, Nofer and Hinz (2015) and Xu et al. (2017) find significant relationships for different time lags. Thereby, we consider this possibility by including one to five lags into empirical models and keep significant variables in the final estimation as reported in Table 3.

To specify the net effect of IMSs on stock returns, we control two daily macro-economic variables, SHIBOR and the exchange rate of Chinese RMB. Meanwhile, we control for a number of calendar anomalies, such as the day of the week effect (French, 1980) and month of the year effect (Al-Hajieh et al., 2011), which have been discussed in the literature. To this end we integrate dummy variables for trading days of Monday, Friday and January. Specifically, the January dummy variable equals one for January 1 to January 31 because of the possible effect of New Year's Day. We also test the December effect and February effect which are demonstrated in western markets and regions using the lunar calendar. The traditional Chinese New Year and holidays are in January 31, 2014 to February 6, 2014. Our empirical tests suggest no significant effects of December, January and February. Therefore, we only report the results of January for brevity.

The following empirical model rooted in our theoretical model Equation (11) is proposed to measure effects of IMSs on stock returns. We estimate Equation (14) with robust standard errors using HAC (Newey-West) due to heteroskedasticity. Meanwhile, if an Autoregressive Conditional Heteroscedasticity (ARCH) effect exists in the residual term, we re-estimate Equation (14) with the variance equation including identified ARCH or Generalized ARCH (GARCH) terms.

$$R_t = \alpha_0 + \sum_j \beta_j R_{t-j} + \sum_k \gamma_{i,k} IMS_{i,t-k} + \sum_m \eta_{1,m} r_{t-m} + \sum_n \eta_{2,n} e_{t-n} + \omega_1 JAN_t + \omega_2 MON_t + \omega_3 FRI_t + \varepsilon_t, \tag{14}$$

where R_t is the return for SHCI, $IMS_{i,t-k}$ denotes orthogonalized sentiment i, and ε_t is the residual error term. JAN_t, MON_t and FRI_t represent dummy variables of January, Monday and Friday, respectively. The control variables of SHCI and exchange rate of RMB are represented by r_{t-m} and e_{t-n}, respectively. The unit root tests based on Augmented Dickey-Fuller unit root test (Dickey & Fuller, 1981, ADF) and Phillips and Perron's test (Phillips & Perron, 1988, PP) suggest that stock returns, IMSs and controlled variables are stationary at the significance level of 1%, thus can be used in the above model. The results of unit root tests are not included for brevity but are available upon request. Particularly, considering correlations between sentiments and macroeconomic variables, the orthogonalized IMSs are used in empirical studies. We compare results regarding different IMSs of six dimensions (anger, disgust, fear, joy, sadness and the total number) with one to five lags and the overall effect based on the first principal component of five specific sentiments (IMS-P). Consequently, approximately 485 models are estimated and the models with best results are summarized in Table 3.

The coefficients of IMSs are significant except the fear sentiment and the incorporating of sentiments improve the adjusted R^2 of corresponding models compared with models without sentiments which is summarized in column 1. The results mean that IMSs' effects on SHCI returns can explain the movements of stock returns better than naive autoregressive models. Specifically, 5-days lagged sentiments are positively related to SHCI return except the disgust sentiment. The results indicate that a larger amount of 5-days lagged sentiments are accompanied with higher SHCI return today. According to Equation (11), there is a mean-reverting component in stock returns under the assumption that the market misperception follows a stationary process. Assets subject to noise trader risk can be underpriced relative to fundamental values. When such misperceptions caused by over pessimistic or over optimistic sentiments fade away, prices approach fundamental values, thus indicating that effects of IMSs on stock returns can revert on a longer term. However, we find no reversal effects for any orthogonalized sentiment. Similarly, we find that the IMS-P has a significantly positive effect on stock returns. This finding is in line with the positive correlations with specific sentiments (Table 2) and the positive effects of five sentiments on stock returns. For other factors, such as calendar effects, the estimates of coefficients identify with results based on specific sentiments. According to Table 3, IMSs can improve the in-sample adjusted R^2 significantly, particularly for joy. The incorporation of joy increases the adjusted R^2 of the naïve autoregressive models with control variables from 0.024 to 0.043. The results suggest that orthogonalized IMSs contain important information of stock returns. Usually, the R^2 of stock return models is smaller than 1% (Nofer & Hinz, 2015), and an increase of 0.5% in the R^2 can significantly improve utilities of risk averse investors (Rapach & Zhou, 2013).

After controlling the weekday effects (Monday and Friday effect), month of the year effect (January effect) and the autocorrelation of stock returns, we find that orthogonalized IMSs are positively related to stock returns, particularly with 5-days lags. The result agrees with the findings of Siganos et al. (2014). Compared with negative sentiments of anger, disgust, fear and sadness, the positive sentiment of joy shows no significant difference in affecting stock returns. Specifically, the signals and values of coefficients are similar with other sentiments. The result runs counter to the conclusion of Chou, Lee, and Ho (2007) which argues a higher risk-

TABLE 4 Effects of holiday IMSs on the SHCI return

	Dependent variable: SHCI return (R_t)							
R_{t-1}	0.097[a]	0.096[b]	0.097[a]	0.097[b]	0.098[a]	0.098[a]	0.098[a]	0.097[a]
$Anger\_H^{\perp}$		0.104						
$Disgust\_H^{\perp}$			0.135[a]					
$Fear\_H^{\perp}$				0.088				
$Joy\_H^{\perp}$					0.079			
$Sadness\_H^{\perp}$						0.095		
$Total\_H^{\perp}$							0.083	
$IMS-P\_H^{\perp}$								0.098
r_t	−0.135[b]	−0.139[c]	−0.140[b]	−0.138[c]	−0.138[b]	−0.138[b]	−0.139[c]	−0.139[c]
e_t	−0.394	−0.426	−0.363	−0.420	−0.396	−0.395	−0.399	−0.403
MON_t	0.135	−0.345	−0.441	−0.229	−0.281	−0.250	−0.357	−0.348
FRI_t	0.137	0.254[b]	0.274	0.223	0.236	0.228	0.256[b]	0.252[b]
JAN_t	−0.195	−0.210	−0.189	−0.209	−0.197	−0.201	−0.202	−0.203
$Cons$	2.825	3.109	2.934	3.049	2.911	2.898	2.940	2.965
Adj. R^2	0.024	0.027	0.030	0.025	0.026	0.026	0.028	0.027

Note: This table summarizes the results of Equation (14) and adjust the residual error with HAC to eliminate the effects of heteroscedasticity. No ARCH effect exists in residual errors. In this table, holiday IMSs are represented by the orthogonalized log values of the numbers of holiday sentiments before the SHCI return in trading t. r_t and e_t represent SHIBOR and exchange rate, respectively. The first column represents the regression results of Equation (14) without any holiday IMS, but with control variables, calendar effects of January effect (JAN_t), Monday effect (MON_t) and Friday effect (FRI_t). $IMS-P\_H^{\perp}$ denotes the first principal component of orthogonalized holiday sentiments. All variables are stationary without unit root, and all models are free of serial correlation problems.
[a]Significance at the 5% level.
[b]Significance at the 10% level.
[c]Significance at the 1% level.

taking tendency for people in good mood compared to those in bad mood.

Meanwhile, we find weak evidence supporting calendar effects in the SHCI. Significant day of the week effect has been found in many markets, such as Standard & Poor's 500 in the U.S., FTSE 30 in the United Kingdom and DAX 30 (German Stock Index) in the German Stock Markets (Alt, Fortin, & Weinberger, 2011). Nevertheless, Carlucci, Júnior, Lima, and Gaio (2014) find no difference between mean returns of each weekday for Bovespa Index in Brazil, Mexican Stock Exchange in Mexico and the DJIA in the U.S. for the period of 2004–2012. Specifically, for the SHCI, a Monday effect is demonstrated in the sample interval with 500 days, and a Friday effect is observed since 2010 on a 1,500 days sample interval. The week of the day effect appears to vary according to different periods. No month of the day effect is demonstrated, thus showing that stock returns in January (the month of the New Year and Chinese Lunar New Year in 2014) are statistically the same as other months. The results are in line with many studies as summarized in Patel and Sewell (2015); but is contrary to Teng and Yang (2018) who suggest a significant Chinese Lunar New Year effect which can mainly attributed to public positive emotions. The coefficients of dummy variables maintain unchanged when different IMSs are added in the model, thus strengthening the robustness of our empirical results.

5.2 | Empirical results considering holiday data

In the above analysis, we estimate the effect of IMSs on stock returns in trading days, but do not consider holiday data. Holiday effects may exist in stock markets, and sentiments in holidays affect stock market performance in the following trading days, but the conclusion is controversial (Swinkels & Vliet, 2012). To test the robustness of our empirical results, we add holiday data in our sample. If day t − 1 is a holiday, the holiday IMS in day t is the sum of sentiments during the holiday period. If the day t − 1 is not a holiday, the holiday IMS in day t is 0. Similarly, to capture the net effect of sentiments in holidays on stock returns, we use the orthogonalized values. Considering that holiday mainly affects stock market performance in the following trading day, we test the

TABLE 5 Effects of holiday IMSs and trading days' IMSs on the SHCI return

	Dependent variable: SHCI return (R_t)							
R_{t-1}	0.097[a]	0.095[b]	0.095[a]	0.100[b]	0.099[b]	0.096[b]	0.097[b]	0.097[b]
$Anger^{\perp}$		0.314[a](5)						
$Anger\_H^{\perp}$		0.093						
$Disgust^{\perp}$			0.183(5)					
$Disgust\_H^{\perp}$			0.124					
$Fear^{\perp}$				0.311[a](5)				
$Fear\_H^{\perp}$				0.075				
$Joy^{\perp}$					0.323[b](5)			
$Joy\_H^{\perp}$					0.064			
$Sadness^{\perp}$						0.265[a](5)		
$Sadness\_H^{\perp}$						0.074		
$Total^{\perp}$							0.376[c](5)	
$Total\_H^{\perp}$							0.071	
$IMS-P^{\perp}$								0.069[a](5)
$IMS-P\_H^{\perp}$								0.085
r_t	−0.135[b]	−0.148[c]	−0.141[b]	−0.144[b]	−0.145[b]	−0.141[c]	−0.145[c]	−0.145[c]
e_t	−0.394	0.297	0.256	0.223	0.385	0.288	0.512	0.427
MON_t	0.135	−0.310	−0.390	−0.162	−0.209	−0.159	−0.289	−0.285
FRI_t	0.137	0.251[b]	0.273[b]	0.237	0.230	0.214	0.258	0.255[b]
JAN_t	−0.195	−0.186	−0.134	−0.168	−0.142	−0.170	−0.153	−0.149
Cons	2.825	−1.323	−1.081	−0.909	−1.891	−1.310	−2.663	−2.141
Adj. R^2	0.024	0.038	0.032	0.036	0.043	0.036	0.043	0.042

Note: This table summarizes the results of Equation (14) and adjust the residual error with HAC to eliminate the effects of heteroscedasticity. No ARCH effect exists in residual errors. In this table, trading days' IMSs ($Anger^{\perp}$, $Disgust^{\perp}$, $Fear^{\perp}$, $Joy^{\perp}$, $Sadness^{\perp}$, $Total^{\perp}$) are orthogonalized values. Holiday IMSs ($Anger\_H^{\perp}$, $Disgust\_H^{\perp}$, $Fear\_H^{\perp}$, $Joy\_H^{\perp}$, $Sadness\_H^{\perp}$, $Total\_H^{\perp}$) are orthogonalized log values of holiday sentiments before the SHCI return in trading *t*. The lags of IMS are in parentheses of the coefficients. r_t and e_t represent SHIBOR and exchange rate, respectively. The first column represents the regression results of Equation (14) without any IMS, but with control variables, calendar effects of January effect (JAN_t), Monday effect (MON_t) and Friday effect (FRI_t). $IMS-P^{\perp}$ and $IMS-P\_H^{\perp}$ denote the first principal component of orthogonalized trading and holiday sentiments. All variables are stationary without unit root, and all models are free of serial correlation problems.

[a]Significance at the 5% level.

[b]Significance at the 10% level.

[c]Significance at the 1% level.

contemporaneous relationship between SHCI return and holiday IMSs in day t. The results in Table 4 show that all coefficients of holiday IMSs are insignificant with one exception that the holiday disgust sentiment is positively related with stock returns. No holiday effects exist for the effect of other sentiments on SHCI returns. The improvement in the adjusted R^2 is minimal, showing that the holiday effect appears to be weak in SSE, the first and largest stock market in China. Otherwise, the addition of holiday sentiments should improve the explanation of stock returns largely.

Since the holiday effect is insignificant for most sentiments, the incorporation of holiday IMSs should exert no effect on the relationship between stock returns and trading days' IMSs. Thus, we combine orthogonalized holiday IMSs and trading days' IMSs in the same model after verifying no significant correlation between them. The results are summarized in Table 5. As expected, orthogonalized holiday IMSs are insignificant for all cases, whereas trading days' IMSs are significantly positive (except for disgust) and the coefficients of them maintain approximately the same as in Table 3. Given the results in Tables 3–5, we can conclude that investors' sentiments induced by macroeconomic factors exert comparatively stable and significant effects on Chinese stock market.

TABLE 6 Forecast accuracy of SHCI returns

	Panel A: Linear forecasting (dynamic forecast)						
	AR1	**Anger**	**Disgust**	**Fear**	**Joy**	**Sadness**	**Total**
RMSE	0.811	0.812	0.802	0.804	0.808	0.691	0.518
MAE	0.633	0.643	0.627	0.636	0.637	0.582	0.461
Theil	0.863	0.822	0.835	0.840	0.811	0.076	0.037
	Panel B: Linear forecasting (static forecast)						
	AR1	**Anger**	**Disgust**	**Fear**	**Joy**	**Sadness**	**Total**
RMSE	0.802	0.803	0.793	0.792	0.798	0.455	0.339
MAE	0.627	0.636	0.621	0.627	0.630	0.338	0.244
Theil	0.834	0.798	0.811	0.816	0.786	0.050	0.024
	Panel C: Nonlinear forecasting (neural network)						
	Naive	**Anger**	**Disgust**	**Fear**	**Joy**	**Sadness**	**Total**
RMSE	1.190	0.963	1.084	1.141	0.873	0.767	0.995
(N)	(11)	(11)	(10)	(9)	(13)	(11)	(11)

Note: This table summarizes the dynamic and static forecast errors of SHCI returns. RMSE denotes the Root Mean Squared Error, MAE denotes the Mean Absolute Error, and Theil is the Theil Inequality Coefficient. AR1 in the first column of Table 4 shows the first order auto-regression results with calendar effects and control variables. The AR model is used because no ARCH effect is found for the residual term. The remaining columns are results with corresponding IMSs. The numbers of hidden neurons are in parentheses, and Naive represents the model without any IMS.

5.3 | Predictability of stock returns

According to Table 3, trading days' IMSs have expansive effects on stock returns. Nevertheless, a good in-sample performance does not imply a superior out-of-sample performance (Wang et al., 2018). Therefore, we estimate the predictability of Equation (14) including static and dynamic forecasting results. Meanwhile, considering that nonlinear relationships are possibly better in capturing the effects of orthogonalized IMSs on stock returns, we further adopt a Neural Network approach to predict stock returns from an out-of-sample perspective. Neural networks are used to decode nonlinear time series data which describe the characteristics of the stock market and predict stock market values (Guresen, Kayakutlu, & Daim, 2011). For example, Bollen et al. (2011) use the Self-organizing Fuzzy Neural Network (SOFNN) to test the ability of tweet moods in predicting stock returns. Particularly, Saurabh and Dey (2020) note that the Artificial Neural Network outperforms many other models such as decision tree, discriminant analysis and support vector machine in using social moods to predict stock returns. Thereby, to assess the contribution of IMSs in predicting stock returns, we compare the performance of a Nonlinear Autogressive with External Input Neural Network model (NARX-NN) that predicts SHCI return on the basis of seven sets of inputs. The naïve set includes open price, close price, highest price, lowest price and trading volumes of SHCI of the past 5 days. The other six sets combine the naïve set with five detailed IMSs and the total number of sentiments. Values of all variables are linearly scaled to [0, 1] to ensure that every input variable is treated with similar importance. To evaluate the NARX-NN model's ability to predict stock returns, we choose April 1, 2013 to June 5, 2014 (75% of the sample data) as the training period, from June 6, 2014 to July 2, 2014 as the validation period which is used to measure network generalization and to halt training when generalization stops improving. The remaining data from July 3, 2014 to October 31, 2014 (20% of the sample data) is the testing period, which provides an independent measure of network performance after training.

The results of prediction accuracy are reported in Table 6. Panels A and B summarize the dynamic and static forecasting accuracy of models estimated in Table 3. Three different indices, that is, the RMSE, MAE and Theil Inequality Coefficient suggest that the incorporation of orthogonalized IMSs help to improve forecast accuracy compared with auto-regression results with calendar effects and control variables. Particularly, sadness and the total number of microblogs with sentiments show comparatively better performance in lowering forecast errors. The regression analysis indicates that disgust sentiment is not significantly related to the SHCI return.

However, it contains predictive information of stock returns. We investigate seven permutations of input variables to the NARX-NN model. The first one is the naïve set, which is the baseline model that has been trained to predict stock returns at time t from the historical values of open price, close price, highest price, lowest price and trading volume at time {t−1, t−2, ..., t−5}. For the commonly used three-layered neural network, the determination of the number of the hidden nodes has no consensus. We follow the rule in determining hidden codes $\sqrt{m+n}+a$, where a denotes an integer in [0, 10], m and n represent the number of input and output variables, respectively. Thus, we train each model with 11 different hidden nodes and select the one with best in-sample forecasting accuracy for our testing. Therefore, the best hidden node for each model is different (as summarized in Panel C, Table 6). Forecasting accuracy is measured in terms of the RMSE of normalized data.

We can draw several conclusions from these results. First, adding IMSs improve prediction accuracy through reducing the RMSE of forecasting results. This confirms the results of regression analysis, meaning that orthogonalized IMSs do contain useful information apart from macroeconomic conditions in improving the predictability of SHCI returns. Combining the results of regression tests, linear forecasting results and NARX-NN, one can find that IMSs are closely related to SHCI returns and contain forecasting information of stock markets. Second, from the perspective of sentiments, joy and sadness appear to be most useful in increasing prediction accuracy. This finding is in line with Saurabh and Dey (2020), in which the public happy mood significantly improves the prediction of the stock return. Overall, IMSs do contain useful information in the out-of-sample prediction of SHCI returns.

Investors' personal sentiments, particularly those caused by private life are not likely to be transferred to the stock market. Our measurement of IMS is based on the effects of changes in the macro-economy on investors' sentiments. Therefore, such sentiment is more closely related to the stock market and more likely to be transferred to the market. As discussed in Bollen et al. (2011), the Opinion Finder's assessment of public mood states in terms of positive and negative mood cannot improve the prediction of DJIA values. By contrast, they demonstrate that the GPOMS dimension labeled “Calm” and the combination of “Calm” and “Happiness” are predictive of the DJIA closing values, but with limited performances. Align with Bollen et al. (2011), we show that detailed sentiments are predictive of stock performances. Furthermore, through controlling macro-economic variables and measuring orthogonalized IMSs, we extract the effects of investors' sentiments on stock returns, which provides additional information that macro-economic factors cannot provide. For example, investors' sentiments can be produced by rumors regarding the macro-economy, which is counter to or exaggerates the fact.

6 | CONCLUSIONS

Many macroeconomic factors affect stock markets; but through unclear channels. In this article, we attempt to produce evidence supporting that the macro-economy can affect stock markets through investors' sentiments. We propose an overlapping generations model based on the noise trader model of De Long et al. (1990) to explain the relationship between the overall investors' sentiments and stock returns. In the model, the market misperception which is mainly driven by pseudo or real signals of changes in the macro-economy determines the equilibrium return of noise traders and sophisticated investors. The market misperception is empirically measured through estimating investors' sentiments that are induced by changes in macroeconomic conditions. We demonstrate net effects of IMSs on stock returns through regression analysis based on orthogonalized sentiments and out-of-sample prediction tests. The empirical results suggest that sentiments are correlated with stock returns, and reversal effects in the relationship between specific IMSs and SHCI returns are insignificant. Furthermore, we demonstrate that idiosyncratic sentiments exert different effects. Robustness tests based upon holiday data and the principal components approve the above conclusions. By comparing with the naïve predicting model without any macroeconomic sentiment, the results suggest that IMSs enhance the out-of-sample predictability of stock returns through linear and nonlinear models, which can significantly improve utilities of risk averse investors. The above results provide reasonable grounds to believe that the macro-economy can affect stock markets through investors' sentiments.

Thousands of macroeconomic, microeconomic and financial factors affect the stock market (Ohno & Ando, 2018). However, limited by information costs, most investors cannot predict the future movements of stock markets as economists. The IMSs measured in this article consider most factors by analyzing microblogs and can to certain extent improve the predictability of stock markets. Although it is difficult for investors to calculate a macroeconomic sentiment as we do, they can indeed surf the Sina Weibo or other SNSs and collect enough information regarding the macro-economy and accordingly correct individual misperceptions to reduce probabilities of losses. Furthermore, given the leading feature of IMSs extracted from SNSs, it is possible and beneficial for the financial regulatory authority to supervise the financial market through tracking the overall information spreading on the Internet, particularly SNSs. Any early sign of market behaviors should be considered carefully to promptly take necessary countermeasures.

The reversal effect does not exist for any sentiment, but may exist for individual stocks. Investors' limited attention is supposed to be a cause of the reversal effect. The number of individual investors dominates that of institutional investors in Chinese financial markets. As assumed in our theoretical model, limited by information asymmetry, individual investors are likely to be impacted by pseudo signals and form misperceptions, thus resulting to drastic changes in stock returns. When investors' sentiments return to normal, the stock price will decline to the fundamental value. By contrast, key opinion leaders on SNSs are supposed to impact public perceptions, thus highlighting possible leading effects of posters with many followers (Nofer & Hinz, 2015). Detailed effects of sentiments on individual stocks and the leading role of SNS influencers such as professional analysts and investment advisors are left for future examination which can be used in portfolio management.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China under Grant 71873014, the Fundamental Research Funds for the Central Universities under Grant 06500106, and the USTB-NTUT Joint Research Program under Grant 06310058.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Yingying Xu https://orcid.org/0000-0001-8253-6930

REFERENCES

Al-Hajieh, H., Redhead, K., & Rodgers, T. (2011). Investor sentiment and calendar anomaly effects: A case study of the impact of Ramadan on Islamic middle eastern markets. *Research in International Business and Finance*, *25*(3), 345–356. https://doi.org/10.1016/j.ribaf.2011.03.004

Alt, R., Fortin, I., & Weinberger, S. (2011). The Monday effect revisited: An alternative testing approach. *Journal of Empirical Finance*, *18*(3), 447–460. https://doi.org/10.1016/j.jempfin.2011.04.002

Bakar, S., & Yi, A. N. C. (2016). The impact of psychological factors on investors' decision making in malaysian stock market: A case of Klang Valley and Pahang. *Procedia Economics and Finance*, *35*, 319–328. https://doi.org/10.1016/S2212-5671(16)00040-X

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*, 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Carabias, J. M. (2018). The real-time information content of macroeconomic news: Implications for firm-level earnings expectations. *Review of Accounting Studies*, *23*(1), 136–166. https://doi.org/10.1007/s11142-017-9436-9

Carlucci, F. V., Júnior, T. P., Lima, F. G., & Gaio, L. E. (2014). The "week-day effect" anomaly in the behavior of stock index returns of Brazil, Mexico and the USA. *Business and Management Review*, *3*(09), 31–38.

Cheng, W., & Lin, J. (2013). Investors' bullish sentiment of social media and stock market indices. *Journal of Management Science*, *27*(5), 111–119. https://doi.org/10.3969/j.issn.1672-0334.2013.05.011

Chou, K. L., Lee, T., & Ho, A. H. (2007). Does mood state change risk taking tendency in older adults? *Psychology and Aging*, *22*(2), 310–318. https://doi.org/10.1037/0882-7974.22.2.310

Christou, C., Cunado, J., Gupta, R., & Hassapis., C (2017). Economic policy uncertainty and stock market returns in PacificRim countries: Evidence based on a Bayesian panel VAR model. *Journal of Multinational Financial Management*, *40*, 92–102. https://doi.org/10.1016/j.mulfin.2017.03.001

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738. https://doi.org/10.1086/261703

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, *49*, 1057–1072.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PLoS One*, *6*(12), e26752. https://doi.org/10.1371/journal.pone.0026752

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*(5596), 1191–1194. https://doi.org/10.1126/science.1076358

Dong, Y. H., Chen, H., Qian, W. N., & Zhou, A. Y. (2015). Micro-blog social moods and Chinese stock market: The influence of emotional valence and arousal on Shanghai composite index volume. *International Journal of Embedded Systems.*, *7*(2), 148–155. https://doi.org/10.1504/IJES.2015.069987

Dräger, L. (2015). Inflation perceptions and expectations in Sweden–are media reports the missing link? *Oxford Bulletin of Economics and Statistics*, *77*(5), 681–700. https://doi.org/10.1111/obes.12078

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, *22*(1), 3–25.

Fan, R., Zhao, J., Chen, Y., & Xu, K. (2014). Anger is more influential than joy: Sentiment correlation in Weibo. *PLoS One*, *9*(10), e110184. https://doi.org/10.1371/journal.pone.0110184

Fisher, K. L., & Statman, M. (2003). Consumer confidence and stock returns. *Journal of Portfolio Management*, *30*(1), 115–127. https://doi.org/10.3905/jpm.2003.319925

French, K. R. (1980). Stock returns and the weekend effect. *Journal of Financial Economics*, *8*(1), 55–69. https://doi.org/10.1016/0304-405X(80)90021-5

Garcia, D. (2013). Sentiment during recessions. *Journal of Finance*, *68*(3), 1267–1300. https://doi.org/10.1111/jofi.12027

Gilbert, E., & Karahalios, K. (2010). Widespread worry and the stock market. International conference on weblogs and social media, Icwsm 2010, Washington, DC, USA, May. DBLP. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/download/1513/1833

Goetzmann, W. N., & Jorion, P. (1993). Testing the predictive power of dividend yields. *The Journal of Finance*, *48*(2), 663–679. https://doi.org/10.1111/j.1540-6261.1993.tb04732.x

Guillory, J., Spiegel, J., Drislane, M., Weiss, B., Donner, W., & Hancock, J. (2011). Upset now? Emotion contagion in distributed groups. In *Sigchi conference on human factors in computing systems* (pp. 745–748). New York, NY: Association for Computing Machinery (ACM). https://doi.org/10.1145/1978942.1979049

Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, *38*(8), 10389–10397. https://doi.org/10.1016/j.eswa.2011.02.068

Hammerschmid, R., & Lohre, H. (2018). Regime shifts and stock return predictability. *International Review of Economics & Finance*, *56*, 138–160. https://doi.org/10.1016/j.iref.2017.10.021

Hong, H., Chen, N., O'Brien, F., & Ryan, J. (2018). Stock return predictability and model instability: Evidence from mainland China and Hong Kong. *The Quarterly Review of Economics and Finance*, *68*, 132–142. https://doi.org/10.1016/j.qref.2017.11.007

Karabulut, Y. (2013). Can Facebook predict stock market activity?. (October 21, 2013). AFA 2013 San Diego Meetings. Retrieved from https://doi.org/10.2139/ssrn.2017099

Keynes, J. M. (1936). *The general theory of employment, interest and money* (Vol. 1936, pp. 3525–3536). England: Royal Economic Society in Macmillan Cambridge University Press. https://doi.org/10.1111/j.1475-4932.1937.tb02999.x

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, *74*(2), 209–235. https://doi.org/10.1016/j.jfineco.2002.11.002

Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, *69*, 14–23. https://doi.org/10.1016/j.knosys.2014.04.022

Liu, X. (2017). Unfolded risk-return trade-offs and links to macroeconomic dynamics. *Journal of Banking & Finance*, *82*, 1–19. https://doi.org/10.1016/j.jbankfin.2017.04.015

Mao, H., Counts, S., & Bollen, J. (2015). Quantifying the effects of online bullishness on international financial markets (No. 9). Paper presented at: ECB Statistics Paper. Retrieved from https://doi.org/10.2866/943096

Neely, C., Rapach, D., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, *60*(7), 1772–1791. https://doi.org/10.1287/mnsc.2013.1838

Nofer, M., & Hinz, O. (2015). Using twitter to predict the stock market. *Business & Information Systems Engineering*, *57*(4), 1–14. https://doi.org/10.1007/s12599-015-0390-4

Oh, C., & Sheng, O. (2011, December). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. In Icis (pp. 1–19). Retrieved from https://aisel.aisnet.org/icis2011/proceedings/knowledge/17

Ohno, S., & Ando, T. (2018). Stock return predictability: A factor-augmented predictive regression system with shrinkage method. *Econometric Reviews*, *37*(1), 29–60. https://doi.org/10.1080/07474938.2014.977086

Patel, N., & Sewell, M. (2015). Calendar anomalies: A survey of the literature. *International Journal of Behavioural Accounting and Finance*, *5*(2), 99–121. https://doi.org/10.1504/IJBAF.2015.072405

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335–346. https://doi.org/10.1093/biomet/75.2.335

Rapach, D., & Zhou, G. (2013). Forecasting stock returns. In *Handbook of economic forecasting* (pp. 327–384). Amsterdam: Elsevier.

Saurabh, S., & Dey, K. (2020). Unraveling the relationship between social moods and the stock market: Evidence from the United Kingdom. *Journal of Behavioral and Experimental Finance*, *26*, 100300. https://doi.org/10.1016/j.jbef.2020.100300

Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*, *107*, 730–743. https://doi.org/10.1016/j.jebo.2014.06.004

Swinkels, L., & Vliet, P. V. (2012). An anatomy of calendar effects. *Journal of Asset Management*, *13*(4), 271–286. https://doi.org/10.1057/jam.2012.9

Teng, C. C., & Yang, J. J. (2018). Chinese lunar new year effect, investor sentiment, and market deregulation. *Finance Research Letters*, *27*, 175–184. https://doi.org/10.1016/j.frl.2018.03.003

Vu, T. T., Chang, S., Ha, Q. T., & Collier, N. (2012). An experiment in integrating sentiment features for tech stock prediction in twitter. Paper presented at: Proceedings of the workshop on information extraction and entity analytics on social media data. The COLING 2012 organizing committee; 2012: pp. 23–38; Mumbai, India.

Wang, Y., Wei, Y., Wu, C., & Yin, L. (2018). Oil and the short-term predictability of stock return volatility. *Journal of Empirical Finance*, *47*, 90–104. https://doi.org/10.1016/j.jempfin.2018.03.002

Xu, Y., Liu, Z., Zhao, J., & Su, C. W. (2017). Weibo sentiments and stock return: A time-frequency view. *PLoS One*, *12*(7), e0180723. https://doi.org/10.1371/journal.pone.0180723

Zhao, J., Dong, L., Wu, J., & Xu, K. (2012, August). Moodlens: An emoticon-based sentiment analysis system for Chinese tweets. Paper presented at: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1528–1531). ACM. Retrieved from https://doi.org/10.1145/2339530.2339772

Zhou, Z., Zhao, J., & Xu, K. (2016, November). Can online emotions predict the stock market in China?. Paper presented at: International Conference on Web Information Systems Engineering (pp. 328–342). Springer, Cham. Retrieved from https://doi.org/10.1007/978-3-319-48740-3_24

How to cite this article: Xu Y, Zhao J. Can sentiments on macroeconomic news explain stock returns? Evidence form social network data. *Int J Fin Econ*. 2020;1–16. https://doi.org/10.1002/ijfe.2260