

RESEARCH ARTICLE

Weibo sentiments and stock return: A time-frequency view

Yingying Xu^{1*}, Zhixin Liu¹, Jichang Zhao¹, Chiwei Su²

1 School of Economics and Management, Beihang University, Beijing, China, **2** Department of Finance, Ocean University of China, Qingdao, Shandong, China

* xuying1221@126.com

Abstract

This study provides new insights into the relationships between social media sentiments and the stock market in China. Based on machine learning, we classify microblogs posted on Sina Weibo, a Twitter's variant in China into five detailed sentiments of anger, disgust, fear, joy, and sadness. Using wavelet analysis, we find close positive linkages between sentiments and the stock return, which have both frequency and time-varying features. Five detailed sentiments are positively related to the stock return for certain periods, particularly since October 2014 at medium to high frequencies of less than ten trading days, when the stock return is undergoing significant fluctuations. Sadness appears to have a closer relationship with the stock return than the other four sentiments. As to the lead-lag relationships, the stock return causes Weibo sentiments rather than reverse for most of the periods with significant linkages. Compared with polarity sentiments (negative vs. positive), detailed sentiments provide more information regarding relationships between Weibo sentiments and the stock market. The stock market exerts positive effects on bullishness and agreement of microblogs. Meanwhile, agreement leads the stock return in-phase at the frequency of approximately 40 trading days, indicating that less disagreement improves certainty about the stock market.



OPEN ACCESS

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

Editor: Wei-Xing Zhou, East China University of Science and Technology, CHINA

Received: January 25, 2017

Accepted: June 20, 2017

Published: July 3, 2017

Copyright: © 2017 Xu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data are available from the figshare database (URLs: <https://dx.doi.org/10.6084/m9.figshare.4559380>).

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Early research on stock market prices, e.g., random walk theory and Efficient Market Hypothesis (EMH) suggest that stock prices are driven by new information that is unpredictable [1, 2]. Later studies, however, present the contrary evidence, supporting that stock market prices can indeed to a certain degree be predicted [3] and casting doubts on the EMH. Furthermore, emotions and mood of financial agents affect investment decisions [4, 5], thus resulting in changes in stock markets. A growing body of research suggests that early indicators can be extracted from online social media (e.g., blogs, Twitter feeds) to predict changes in the stock market [6]. In contrast, [7] identifies that sentiments regarding stock markets have no discernible effect on stock prices.

Research that examines whether and how microblogging messages are related to stock indicators appears to present promising results in the United States, Eastern Europe, and Germany

[8, 9]. For example, [10] finds that certain mood states on Twitter (www.twitter.com) are related to the closing values of the Dow Jones Industrial Average (DJIA). The study [6] reports a high correlation between stock prices and Twitter sentiments and Granger causalities running from Tweet sentiments to stock prices. [11] explores data from StockTwits, a microblogging platform exclusively dedicated to the stock market, and documents that the posting volume can improve the forecast of trading volume in the stock market. Experimental results in [12] demonstrate that sentiments of Tweets with cash-tags, which indicate that corresponding Tweets are related to stocks, help improve the prediction of stock returns markedly. Although certain evidence is provided [13, 14, 15], to the best of our knowledge, the relationship between microblogging messages and stock markets remains empirically understudied in China. In contrast to most emerged markets, the marketing policy intervention in China will introduce non-market factors that may disturb the stock market. Moreover, those possible interventions can be leaked through social media and then affect investors' sentiments and decisions. Given that there is a significant quantity of individual investors in China, online news is likely to impact the stock market as indicated by the conventional behavioral finance theory. Indeed, [16] demonstrates that the establishment of Sina Weibo, one of the most popular social networking platform, as an alternative information interaction channel has changed the overall stock market behavior. The evolution of the Internet has made it easier to obtain interactive data and has created new channels for information interaction. Therefore, the relationship between sentiments extracted from Sina Weibo and the stock market performance may provide certain evidence regarding the prediction of stock prices [17]. Moreover, with the wild evolutions of social media and stock markets, relationships between sentiments of microblogs and stock markets are not likely to maintain unchanged in recent years. This paper makes the following contributions to the literature.

First, in contrast to related previous studies that assume a constant relationship between sentiments of online social media and the stock market, this study considers possible changes in both time and frequency domains in the relationship. Previous literature, in general, refers to methods of the Pearson correlation and Granger causality tests that test unchangeable relationships between series. Thus, those results are susceptible to misleading conclusions in the presence of changes in the relationships. Investigations based on different time horizons may come to contrary conclusions. This paper adopts a Wavelet Coherency Analysis based on the continuous wavelet transformation to revisit the relationship between sentiments extracted from Sina Weibo (www.weibo.com) and the stock market.

Second, data based on Chinese microblogs from Sina Weibo, a Twitter-like service, help weaken the sample-bias. Most previous studies focus on microblog postings from Twitter that are in English. As Twitter's user base becomes increasingly international, the bias induced by geographical and cultural sampling errors is unavoidable [10]. Sina Weibo launched in China in 2009 and accumulated over 500 million users in fewer than four years. Most users locate in China and post Chinese microblogs, thus weakening the bias of sampling errors.

Finally, most previous studies categorize microblogs into positive and negative sentiments, ignoring more detailed dimensions. The oversimplification of the emotion classification makes it difficult to disclose detailed correlations of specific sentiments. Negative emotions such as anger, disgust or sadness are more applicable in real world scenarios, i.e., abnormal event detection or emergency tracking [18]. Thus, figuring out the relationships between detailed sentiments and stock markets may shed light on understanding how abnormal events affect financial markets through the network.

This paper collects 157,351 microblogs regarding Chinese stock markets from June 2014 to December 2014, and classifies the sentiment of an individual microblog into five categories, including anger, disgust, fear, joy, and sadness. To compare with previous research, we also

categorize emotions into negative and positive and construct two indices of Bullishness and Agreement. We focus on relationships between various sentiments and the stock market and examine how such relationships change over time and across different scales (frequencies). To do so, we utilize the continuous wavelet analysis which can localize correlations between series and evolution in time and across scales. Both time and frequency are important for stock price movements considering that the Chinese stock market has undergone a wild evolution in recent years. It is reasonable to believe that relationships between sentiments of microblogs and the stock market have changed. Additionally, the frequency domain viewpoint provides an opportunity to distinguish between short- and long-term correlations. We show that relationships between sentiments of microblogs and the Chinese stock market are not always significant, particularly at high frequencies. The most significant relationship appears surrounding ten trading days, and sadness has a closer relationship with the stock return than the other four sentiments. Compared with polarity sentiments (positive and negative), detailed sentiments show more persistent effects. Bullishness has a loose and unstable relationship with the stock market. Regarding the lead-lag relationship, the agreement of investors leads the stock market in-phase over the long run. However, we find that the stock market leads detailed and polarity sentiments rather than the reverse. Our results suggest that relationships between Weibo sentiments and the stock market change over time and across frequencies and detailed sentiments contain more information regarding the stock market than polarity sentiments. Furthermore, causalities running from agreement to the stock return indicate contrary evidence regarding the EMH.

Literature review

Over the past decade, significant progress has been made in sentiment tracking techniques that extract indicators of public mood directly from social media content, such as Internet stock message boards like Yahoo! Finance, in particular large-scale Twitter feeds and microblogs [18, 19]. [20] is the first to investigate the relationship between internet stock message boards and stock market activities. Using a sample of over 3,000 stocks listed on Yahoo! message boards, the study finds that changes in daily posting volume are associated with earnings-announcement events and daily changes in stock trading volume and returns. Consequently, the volume of messages predicts changes in next day stock trading volume and returns. [21] examines the relationship between the Internet message board activity and abnormal stock returns and trading volume based on the RagingBull.com discussion forum. They report that changes in investor opinions correlate with abnormal industry-adjusted returns and abnormally high trading volume on days with eccentrically high message activity. No such linkage is found between the message board activity and industry-adjusted returns or the trading volume, which is consistent with the EMH. [22] categorizes stock recommendations (buy, sell, and neutral) through a database under Google's ownership, and suggests that these recommendations are overwhelmingly positive. Nevertheless, no evidence supporting that the recommendations have new information is found, indicating no significant linkage between internet stock messages and stock market activities. [23] establishes that stock microblog sentiments do have predictive power for simple and market adjusted returns using microblog postings from StockTwits and Yahoo! Finance, respectively.

The above studies focus on readily available quantitative information such as the message volume and user ratings and do not capture the information content and sentiments of microblog messages. As emphasized in [19], newsworthiness may be linked to tones, positive or negative. They identify a gatekeeping function, implying that newsworthiness may be associated with negativity across a wide range of subjects. The finding is buttressed by the literature on

individuals' disproportionate attentiveness to negative versus positive information, across many subjects, e.g., [24, 25]. These literature suggest that negative information is in certain situations viewed as being more important than equally positive information. Similarly, [26] finds that negative words contain more information than positive words.

Thus, contents and sentiments implied in messages from social media are captured by automated classifiers and are used to predict stock market activities. Most studies appeal to the binary sentiment (positive vs. negative) analysis. For instance, by employing an online sentiment classifier called Twitter Sentiment Tool (TST), [27] detects sentiments of Tweets as positive and negative. Based on the Granger causality analysis, they use sentiments and the consumer confidence in the product to address the problem of predicting daily up and down movements in individual tech stock prices. Their results indicate that causalities running from Tweet sentiments to stock prices exist. The study [13] assigns the respective label (positive or negative) for each Tweet and demonstrates that it holds significant predictive power for subsequent stock market movements. [28] uses a supervised machine learning approach to train a sentiment classifier, and categorizes a Tweet into one of the given classes (positive or negative sentiment). By using the Granger causality test, they show that positive and negative sentiments indicate stock price movements a few days in advance. Using a regression analysis, [29] demonstrates a significant relationship between a specific firm's stock return and the cumulative emotional valence (positive or negative) of Tweets regarding the firm. [30] finds a relatively low Pearson correlation and a Granger causality between the sentiment regarding the 30 stock companies (positive or negative) posted on Twitter and the corresponding Dow Jones Industrial Average (DJIA) index over the entire period. However, the result based on "event-study" shows that the sentiment polarity during the peaks of Twitter volume implies the direction of cumulative abnormal returns. Thus, polarity sentiments appear to contain useful information regarding stock activities.

Sentiment analysis tools used in the literature adhere to a binary distinction between positive and negative sentiments. Nevertheless, rich and multi-dimensional structures of human moods were ignored [10]. The Google-Profile of Mood States (GPOMS), a mood analysis tool employed in [10] captures human mood states regarding six different mood dimensions, namely alert, calm, happy, kind, sure, and vital. They establish that the mood time series identified by GPOMS respond to significant socio-cultural events such as the Presidential Election and Thanksgiving. According to the results of Granger causality tests and Self-Organizing Fuzzy Neural Network, they conclude that the calm sentiment identified by GPOMS is predictive of the DJIA. By contrast, general levels of positive and negative sentiments measured by another mood analysis tool of Opinion Finder cannot predict the DJIA. In other words, if there is a great deal of "calm" emotion in Tweets on a given day, the DJIA tends to rise over the following two to five days.

Whereas these studies have investigated the relationships between online social media sentiments and stock markets, most are based on data from emerged markets such as the United States and Europe. With the establishment of Sina Weibo, the microblogging web-site dominator in China [31], more focus is put on the Chinese market. [32] shows positive relationships between investors' bullish sentiment indices from Sina Weibo and stock market indices using Granger causality tests and impulse responses. [33] uses a sentiment analysis technology to automatically classify reviews from Sina Finance, a typical financial website related to Sina Weibo, as positive or negative, and then identifies investor sentiment as either bullish or bearish. Their empirical results suggest solid correlations between stock price volatility trends and stock forum sentiments. [34] shows that the bullish and bearish stock market conditions extracted from Sina Weibo articles have strong correlations with the Shanghai Composite Index (SHCI) particularly when the data set has distinguishing characteristics. [31] documents

that stocks with high original and re-post microblogs are often accompanied by low average volumes per transaction, implying a close linkage between microblogs and the stock market. Using Harvard psychological dictionary and Loughran—McDonald financial sentiment dictionary, [35] constructs a sentiment space and suggests that sentiments from news articles improve the prediction of Hong Kong Stock Exchange prices. Nevertheless, simply focusing on positive and negative dimensions cannot bring useful predictions. [36] obtains five sentiments including anger, disgust, fear, happiness, and sadness by analyzing the text content of daily microblogs from Sina Weibo using an emotion lexicon. Results of the Granger causality analysis and linear regression suggest that the sadness sentiment improves the predictive accuracy of trading volume. In the study of [37], the results based on correlation analysis and causality test show that the stock market in China can be competently predicted by Weibo emotions, e.g., disgust, fear, joy, and sadness.

The linkage between sentiments of online social media and financial markets is assumed to be constant in the above investigations, which is not likely to be true. The following studies provide evidence that contradicts the constant relationship. [38] reveals that 71% of the messages posted on Twitter do not travel any farther than the authors' timeline, thus indicating that most of the information is hardly propagated through the network. Based on machine learning techniques and the inference of time-dependent diffusion probabilities from a multi-dimensional analysis of individual behaviors, [39] illustrates two peaks in the transmissions of Tweets. There is a quick outbreak of the information at the beginning, and the transmission appears to die out before another peak of activity occurs. Furthermore, [40] finds that approximately 73% of topics have a single active period and 15% and 5% of topics have two and three active periods, respectively. Very few have more than three active periods. Consequently, sentiments generated by posts on social media are more likely to have structural changes and the linkage between sentiments and stock markets is time-varying. [32] provides certain direct evidence, in which the results of impulse response show that short-term effects of investors' bullish sentiment indices of posts on Sina Weibo on stock returns are significant. In a related study, [41] applies a continuous wavelets framework and finds that the Bitcoin price is positively related to search queries on Google and searched words on Wikipedia. Furthermore, the relationship changes over time and frequency. Therefore, we explicitly investigate the time-varying and frequency-related linkage between sentiments of stock related microblogs on Sina Weibo and stock market activities in China using the wavelets analysis.

Methodology

Sentiment classification

We accumulate 157,351 microblogs with keywords relating to Chinese stock markets from June 2014 to December 2014 by crawling posts on Sina Weibo through its open Application Programming Interfaces (APIs) under the authority granted by Sina Weibo. The content of microblogs is recorded in the form of short text which is limited to 140 characters. Many approaches have been proposed to mine sentiments from texts in recent years. For example, the lexicon-based method is widely used. [42] uses human evaluations of the emotional content of individual words within a given text to quantify the overall score of happiness levels on large-scale texts. By employing Amazon Mechanical Turk, [43] scores over 10,000 English words on an integer scale from one to nine, where one represents sadness, and nine represents happiness. Using data from millions of public Twitter messages, [44] identifies individual-level mood rhythms using Linguistic Inquiry and Word Count (LIWC). Another widely used method is machine learning based solution. [45] runs classifiers trained on emoticon data against a test set of Tweets (which may or may not have emoticons in them) which are

classified as either positive or negative. In contrast to most works which categorize emotions into negative and positive, [18] employs emoticons for the generation of sentiment labels for microblogs on Sina Weibo and builds an incremental learning Naive Bayes classifier for the categorization of four types of sentiments: angry, disgust, joy, and sadness.

This paper adopts the method in [38, 46] to classify sentiments of microblogs from Sina Weibo. Numerous training samples are required to handle the extremely short text in microblogs. Considering that both smiley and emoticons are strongly related with typical sentiment words and are viewed as convincing indicators of emotions [47], we treat these emoticons as sentiment labels. The most commonly used 95 emoticons are selected for the labeling stage. Several experts label their emotions manually based on the images of emoticons and approximately 50 frequent words occurring together with emoticons. We find that most of the emoticons are labeled by five sentiments, including anger, disgust, fear, joy, and sadness. For other emotions such as surprise, insufficient votes are found. Thus, we classify the sentiments of microblogs as five classes.

Approximately 10,000 microblogs with valid emoticons are extracted and labeled. Using the labeled emoticons as a training corpus C_{tr} , we build a fast Bayesian classifier to mine the sentiment of microblogs without emoticons. For each microblog t in C_{tr} , we convert it into a sequence of words $\{W_i\}$, where W_i is a word, and i is its position in t . The prior probability of the world W_i belonging to the sentiment category m is provided by:

$$P(W_i/m) = \frac{n^m(W_i) + 1}{\sum_q [n^m(W_q) + 1]} \tag{1}$$

where $m = 1, 2, 3, 4, 5$, $n^m(W_i)$ denotes the times that W_i appears in all microblogs in the category m , and the Laplace smoothing is used to avoid the problem of zero probability. For an unlabeled microblog t with word sequence $\{W_i\}$, its category can be obtained as

$$m^*(t) = \arg \max_m P(m) \prod_i P(W_i/m) \tag{2}$$

where $P(m)$ is the prior probability of m . We set $P(m) = 0.2$ to guarantee the overall recall of the classifier. Using Eq (2), we can classify microblogs without emoticons into five sentiments. [18] trains a fast Naive Bayes classifier on data from Sina Weibo for emotion classification, and the system named MoodLens whose vital part is the emotion classifier is now available online for temporal and spatial sentiment patten discovery.

Wavelet analysis

The wavelet transform is a commonly used signal conversion and processing method, which can be traced back to the Fourier analysis. Wavelets are localized in both time and frequency domains and allow for analysis of time-frequency dependencies between two time series. In the case of the wavelet transform, the time resolution is intrinsically adjusted to a frequency with the window width narrowing when focusing on high frequencies whereas widening when assessing low frequencies. Furthermore, wavelets are useful for processing non-stationary time series and extracting information of time domain localization [48]. Compared with commonly used approaches of the Pearson Correlation and the Granger Causality test [12, 13], the wavelet analysis can identify the time-varying relationships between microblog sentiments and stock returns, and it considers the frequency domain [41].

Wavelet Transformation is a process of decomposition and superposition of information. By adding a series of basis wavelet which is obtained by translation and dilation of the so-called mother wavelet, the original time series is transformed into a two-dimensional plane of time-frequency. The continuous wavelet transform (CWT) $W_x(\tau, s)$ is obtained by projecting the

mother wavelet $\psi(t)$ into examined time series $X(T)$. The mother wavelet is defined as:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \tag{3}$$

where τ denotes the location parameter which measures the time position of the mother wavelet, and s is the scale parameter which indicates how the wavelet is stretched or dilated. If $s < 1$, the wavelet is compressed and captures rapid changes at high frequencies. If $s > 1$, the wavelet is stretched and captures slow changes at low frequencies. Given a mother wavelet, the CWT is defined as:

$$W_x(\tau, s) = \int_{-\infty}^{+\infty} x(t) \psi_{\tau,s}^*(t) dt \tag{4}$$

where $\psi_{\tau,s}^*(t)$ represents the complex conjugate of the basis wavelet $\psi_{\tau,s}(t)$. Consequently, the wavelet transform decomposes a time series $x(t)$ regarding certain basis wavelets that are obtained by translation and dilation of the mother wavelet $\psi(t)$. The most commonly used mother wavelet in analyzing information on both amplitude and phase is the Morlet wavelet [49], which is a complex sine wave within a Gaussian envelope:

$$\psi_{\omega_0}(t) = \pi^{-1/4} (e^{i\omega_0 t} - e^{-\omega_0^2/2}) e^{-t^2/2} \tag{5}$$

In Eq (5), $\pi^{-1/4}$ ensures that the wavelet function has unit energy, and $e^{-\omega_0^2/2}$ ensures the admissibility condition of a mother wavelet. If $\omega_0 > 5$, $e^{-\omega_0^2/2}$ can be ignored [50]. In this case, the Morlet wavelet is simplified as follows:

$$\psi(t) = \pi^{-1/4} e^{i\omega_0 t} e^{-t^2/2} \tag{6}$$

The parameter ω_0 in Eqs (5) and (6) denotes the wavenumber, which measures the number of oscillations within the Gaussian envelope. There are two conflicting objectives in choosing ω_0 : frequency localization and time localization. A greater ω_0 is accompanied by a better frequency localization but a poorer time localization, and vice versa. As noted in [40], ω_0 is usually set to six to guarantee a good balance between time and frequency resolution.

The wavelet power spectrum of a time series $x(t)$ is defined as the modulus square of the CWT, i.e. $|W_x(\tau, s)|^2$, which recovers the relative contribution at each time and each scale to the time series variance. The wavelet power spectrum can be integrated across τ and s to recover the total variance in the series:

$$\sigma_x^2 = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |W_x(\tau, s)|^2 \frac{d\tau ds}{s^2}, \text{ with } 0 < C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty \tag{7}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$. The cross-wavelet transform of two time series $x(t)$ and $y(t)$ is defined as $W_{xy}(\tau, s) = W_x(\tau, s)W_y^*(\tau, s)$ [51]. The cross-wavelet spectrum is correspondingly defined as $|W_{xy}(\tau, s)|^2 = |W_x(\tau, s)|^2 |W_y(\tau, s)|^2$, implying the local co-movement between $x(t)$ and $y(t)$.

The wavelet coherency coefficient measures the local strength of the relationship between two series over time and across frequencies.

$$R^2(\tau, s) = \frac{|S(s^{-1}W_{xy}(\tau, s))|^2}{S(s^{-1}|W_x(\tau, s)|^2)S(s^{-1}|W_y(\tau, s)|^2)}, \text{ with } R^2(\tau, s) \in [0, 1] \tag{8}$$

where $S(\cdot)$ is the smooth factor in time and scale normalization. Without smoothing, the coherency is identically one at all scales and times. Smoothing is achieved by convolution in

time and scale. More details please refer to [48]. s^{-1} converts to an energy density [52]. The value of $R^2(\tau, s)$ should be within the interval [0, 1]. The greater value of wavelet coherency coefficient indicates the stronger comovement between $x(t)$ and $y(t)$. Based on a large number of Monte Carlo simulations, [53] documents that the wavelet power spectrum asymptotically follows the χ^2 distribution with two degrees of freedom. As a consequence, we can test the significance of the correlation by testing the null and alternative hypothesis as follows:

$$H_0 : \left[\frac{|W_x(\tau, s) W_y^*(\tau, s)|}{\sigma_x \sigma_y} \right] \leq \frac{3.999}{2} \sqrt{P_f^x P_f^y} \tag{9}$$

$$H_1 : \left[\frac{|W_x(\tau, s) W_y^*(\tau, s)|}{\sigma_x \sigma_y} \right] > \frac{3.999}{2} \sqrt{P_f^x P_f^y} \tag{10}$$

where P_f^x and P_f^y are the background spectra of $x(t)$ and $y(t)$ under the χ^2 distribution. If the null hypothesis H_0 is rejected, we interpret the result as $x(t)$ and $y(t)$ are correlated with each other under the significance of 5%. If the null hypothesis cannot be rejected, we tend to believe that the correlation between the two indices are insignificant.

To investigate the lead-lag relationships, we use the wavelet phase difference between $x(t)$ and $y(t)$ [53]. The wavelet phase difference is defined as the ratio of the imaginary component of the cross-wavelet transform $W_{xy}(\tau, s)$ to the real part:

$$\phi(\tau, s) = \arctan \left(\frac{\mathcal{I}(s^{-1} W_{xy}(\tau, s))}{\mathcal{R}(s^{-1} W_{xy}(\tau, s))} \right), \text{ with } \phi(\tau, s) \in [-\pi, \pi] \tag{11}$$

Graphically, the phase difference is represented by an arrow. If the arrow points to the right (left), the series are in the in-phase (anti-phase), meaning that they are positively (negatively) correlated. If the arrow points down (up), $x(t)$ leads $y(t)$ by $\pi/2$ (vice versa). Usually, the relationship is a combination of the two. For example, if the arrow points to the northeast, the series are positively correlated, and $y(t)$ leads $x(t)$. Note that the interpretation of phase relationships is partially dependent on specific expectations regarding the relationship because a leading relationship in the in-phase can be a lagging relationship in the anti-phase.

Data

We have collected 157,351 stock-related microblogs of 12 to 3,292 daily postings; this represents an average of 1,299 posts per trading day with a standard deviation of 623 messages. Using the classification method explained previously, we classify microblogs into five categories, i.e., anger, disgust, fear, joy, and sadness. Furthermore, as a comparison, we count the numbers of posts with the sentiment of joy as positive microblogs and the remaining as negative microblogs for each day of the time series. As suggested in [54], it is the natural starting place to test whether the level of microblogs or the bullishness of them affect stock markets. The second issue is whether a greater disagreement is associated with the stock market. The financial theory provides two distinct perspectives on disagreement. Whereas [55] and others hold the hypothesis that disagreement induces trading, the “no-trade theorem” [56] argues that disagreement leads to a revision of stock prices. Following the literature [6, 11, 57], we define the proxy of Bullishness for each day as:

$$Bullishness_t = \ln \left(\frac{1 + M_t^{Positive}}{1 + M_t^{Negative}} \right) \tag{12}$$

where $M_t^{Positive}$ and $M_t^{Negative}$ denote the number of positive or negative microblogs on a particular day t . The index of Bullishness measures the share of surplus positive signals and gives more weight to a greater number of microblogs in a specific sentiment (positive or negative). The agreement among positive and negative microblogs is represented by:

$$Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{Positive} - M_t^{Negative}}{M_t^{Positive} + M_t^{Negative}} \right)^2} \quad (13)$$

The value of $Agreement_t$ is greater if more agents share the same sentiment. If all microblogs regarding the stock market are of the same sentiment, Agreement would be one in that case. The Shanghai Composite Index (SHCI) denotes the average stock price of Shanghai Stock Market. Thus, we use the stock return based on the closing price of SHCI as the index of the stock market. The data are available in Wind database.

We test the correlation between microblog sentiments and the stock return in [Table 1](#). According to the Pearson Test, which tests the linear relationship between two variables, the numbers of microblogs with detailed sentiments exhibit different linkages with the stock market. Fear has a positive linear relationship with the return of SHCI, indicating that more fear sentiments are likely to increase overall stock returns of the Chinese market. However, no linear relationship is found in other sentiments. The non-linear relationship tests of Kendall's τ_b test and Spearman correlation test suggest that fear and sadness sentiments and Bullishness are positively related to the return of SHCI. The agreement of posts on Sina Weibo shows a negative linkage with the stock market in three tests. The positive relationship between Bullishness and the stock market indicates that more microblogs sharing the positive sentiment are accompanied by a greater SHCI return. The result agrees with the behavioral economics hypothesis suggesting that when people are happy and optimistic, they are more likely to increase investment, which in turn improves the stock market performance [13]. On the other hand, the results of fear and sadness sentiments imply a contrary conclusion, suggesting that more such pessimistic sentiments are accompanied by better stock performance. Furthermore, the correlation between Agreement and SHCI return is significantly negative, indicating that if agents are more concentrated in one sentiment, the stock market is likely to have a lower return. The results of linear and nonlinear correlation tests support close linkages between microblog sentiments and the stock market for the full sample period, in line with previous literature [58, 59, 60]. However, the results provide no information regarding the lead-lag relationships between social media sentiments and the stock market.

The proposed visual presentation of sentiment time series and stock return can be seen in [Figs 1](#) and [2](#). Peaks show the days when people intensively posted microblogs regarding the stock market. Two prominent negative peaks can be observed in October and December, and one positive peak shows up in December. This type of visualization can be used as a tool for easier and faster overview analysis of general observation of trends. It is interesting to observe that the correlations between microblog sentiments and the stock return are not fixed. For example, the sentiments are comparatively stable in the first half sample and are mostly negative. As sentiment reversed its polarity in September, remarkable fluctuations appeared in the second half of the sample. Analogously, the return of SHCI fluctuated more widely during October to December. If we focus on the contemporary movement between sentiments and the stock return, positive linkages are likely to be observed. For instance, during September 29th to October 17th, the stock return shares the same trends with both positive and negative sentiments. However, in the following month, the stock return moves in the opposite direction of negative sentiment. Meanwhile, positive sentiment coordinates the movement of stock

Table 1. Correlations between microblog sentiments and the stock return.

	Stock Return		
	Pearson test	Kendall's τ_b test	Spearman test
<i>Anger</i>	0.112	0.025	0.033
<i>Disgust</i>	0.087	0.086	0.116
<i>Fear</i>	0.198**	0.106*	0.150*
<i>Joy (Positive)</i>	0.068	0.000	0.006
<i>Sadness</i>	0.114	0.109*	0.163**
<i>Negative</i>	0.093	0.055	0.075
<i>Agreement</i>	-0.202**	-0.112**	-0.173**
<i>Bullishness</i>	0.127	0.106*	0.163**

Notes:
 ** and * denote significance at the 5% and 10% level, respectively. These tests are used by Stata software. The results of positive sentiment are the same as joy.

<https://doi.org/10.1371/journal.pone.0180723.t001>

return. Thus, the relationship between sentiments and the stock market is not constant and appears to be different across frequencies.

More specific changes in sentiments can be observed from Fig 2. A significant decrease in anger can be found surrounding November 21st, when the central bank of China announced a cut in the interest rate for the first time in the past 28 months, thus inducing a steep increase in the capital market. At the end of October, when half stocks encountered sharp decreases, and the other half increased significantly, microblogs with disgust and joy sentiments reached comparatively high levels and accounted for approximately 70% of all posts. When the daily trading volume exceeded 1.07 trillion Yuan on December 5th, most investors were excited and

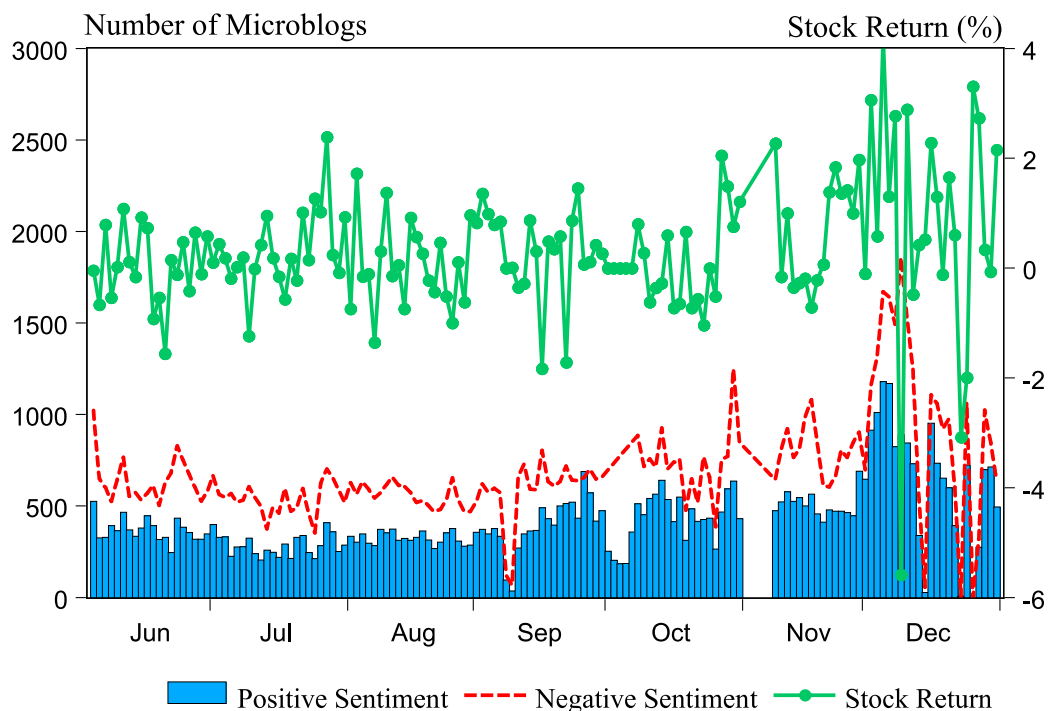


Fig 1. Positive and negative microblogs and the SHCI return.

<https://doi.org/10.1371/journal.pone.0180723.g001>

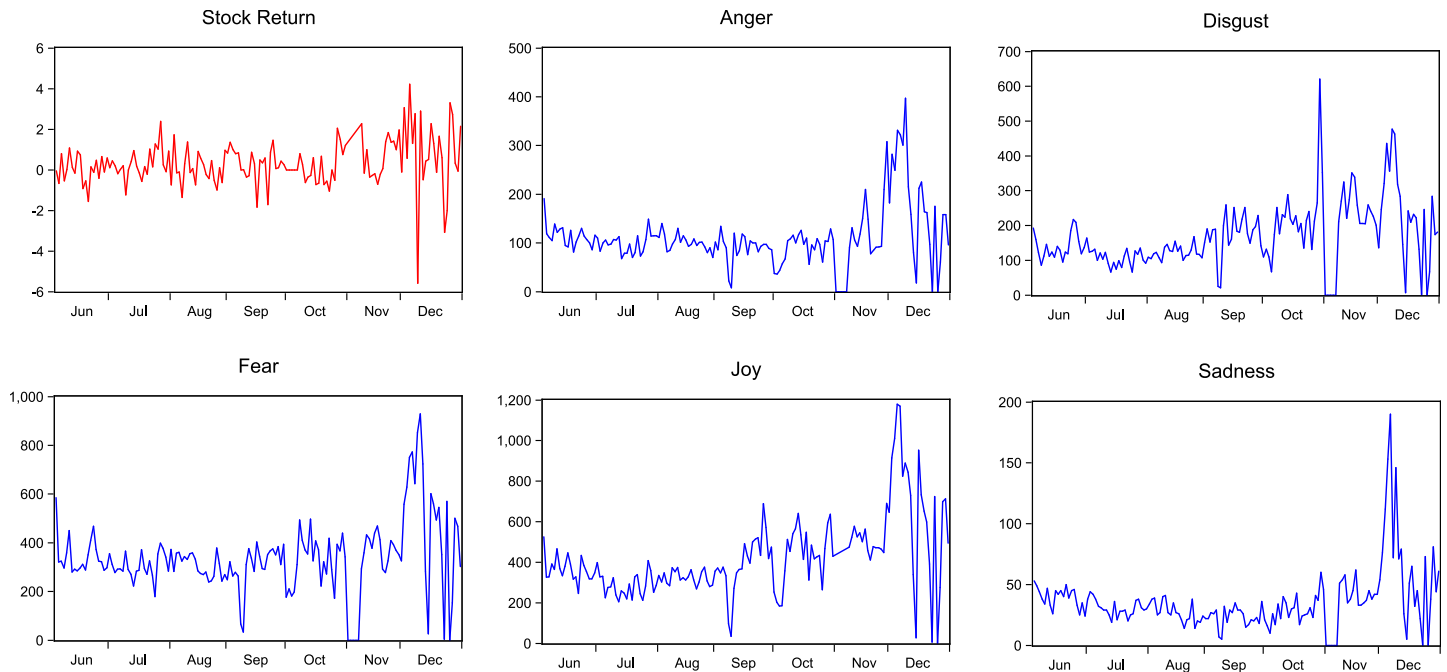


Fig 2. Microblogs with anger, disgust, fear, joy, sadness sentiments and the SHCI return.

<https://doi.org/10.1371/journal.pone.0180723.g002>

posted approximately 1,169 microblogs with joy sentiment. By contrast, the abnormal performance of the stock market triggered certain investors' concern and fear about the future market. Many agents were worried about the short-term correction of the stock market and its departure from the real economy. Consequently, the sentiment of fear appeared a high level. Subsequently, this concern appears to come true with the stock return decreased to negative on December 9th and disgust and fear reached peaks. On the evening of December 9th, the Securities Registration of China announced the measures about strengthening risk management of bond repurchase, which was viewed as a black swan of the bond market. Consequently, the SHCI receded approximately 9% from the peak among 49 months and decreased 5.4% in one day. Meanwhile, the negative sentiment posted on Sina Weibo reached a peak. During December 3rd to December 15th, investors were suffered from the fluctuating stock market and showed peaks and troughs in all sentiments. The changes suggest that sentiments extracted from Sina Weibo are closely related to the return of SHCI and captures significant breaks of the Chinese stock market.

Empirical results

Morlet wavelet power spectra

Fig 3 presents the Morlet wavelet power spectra of the return of SHCI, microblog sentiments, Bullishness, and Agreement. The horizontal axis denotes time, and the vertical axis represents different scales (or frequency) (i.e. the wavelet transform periods). We categorize time scales of fewer than five trading days as short-term, those between 5–20 trading days as medium-term, and those more than 20 trading days as long-term. The black contour areas represent significance at the 95% confidence interval. High power value of a certain variable is represented by highlighted red area, suggesting that the spectral energy is high, and vice versa. The high power areas indicate the impulse from events, showing that the variable fluctuates significantly.

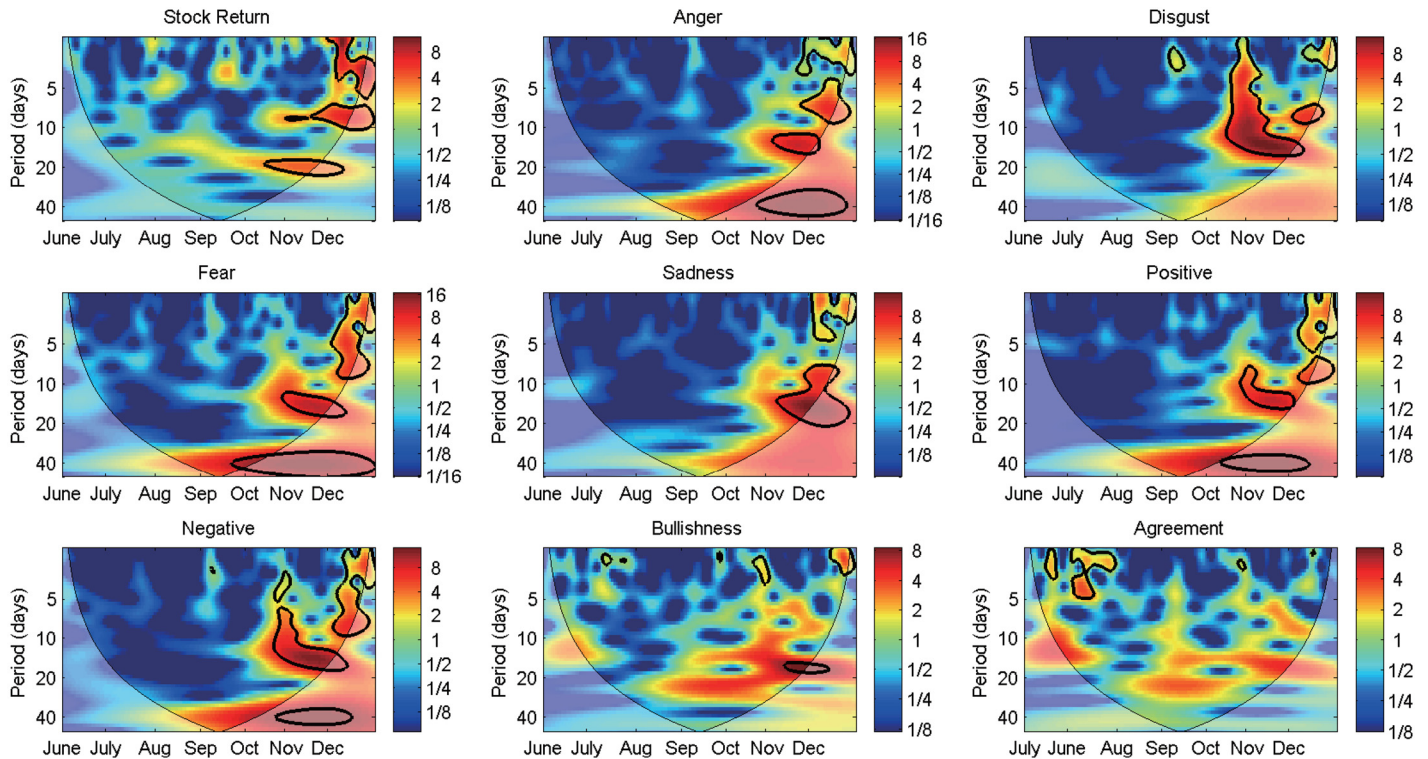


Fig 3. Morlet wavelet power spectra of sentiments and the SHCI return.

<https://doi.org/10.1371/journal.pone.0180723.g003>

The cones formed with bold black lines bending upward denote the Cones of Influence (COI), representing the boundary conditions of the Morlet wavelet transform. The CWT assumes that the data are circulatory, thus making certain bias at the beginning and end of the sample period in addressing finite-length time series. Thereby, outer regions of the COI indicate that fluctuations in variables are sensitive to the edge effect.

The wavelet spectra provide a first assessment of the individual behavior of data in a time and frequency varying framework. The wavelet power spectrum of stock return shows no significant fluctuation for most of the sample period before November. During November to December, the return of SHCI records three regions of high power spectra at frequencies of fewer than five and 5–20 trading days, suggesting that the spectra are different among distinct time-frequency spaces. As with the stock return, five sentiments show significant fluctuations among both time and frequency domains. Most high power spectrum regions appear surrounding November, and detailed sentiments fluctuate significantly over medium to high frequencies, matching the intense short to mid run cyclical oscillations in the time series during that period. No significant fluctuation is found at low frequencies except for fear. Negative sentiment shares a similar pattern with anger and disgust and appears to merge idiosyncratic characteristics of more detailed sentiments. The movements of Bullishness and Agreement are moderately different from other sentiments with the fluctuations in both indices being less persistent and less pervasive. For instance, Agreement fluctuates significantly at frequencies of fewer than five trading days in August. Bullishness is comparatively stable over the entire sample period. Similar fluctuations in five sentiments and the stock market are captured in November and December, but distinct characteristics exist.

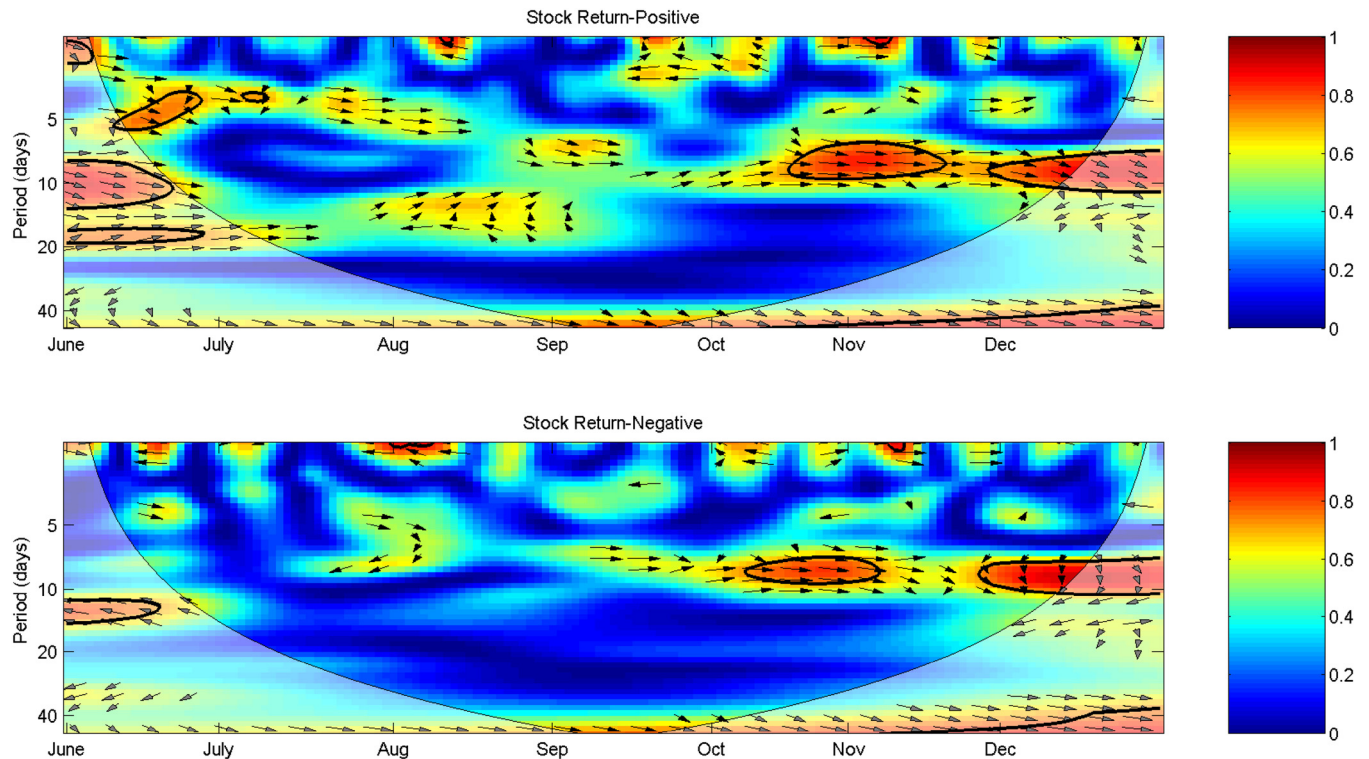


Fig 4. Morlet wavelet transform cross-spectra and coherence spectra of polarity sentiments and the SHCI return.

<https://doi.org/10.1371/journal.pone.0180723.g004>

Coherence analysis

The coherence and phase relationship of microblog sentiments and the stock return illustrate the co-movement and lead-lag relationships between indices. Fig 4 shows the Morlet wavelet transform cross-spectra and coherence spectra of polarity sentiments and the SHCI return. In Fig 4, the color code for coherency ranges from blue corresponding to low coherency (close to zero) to red corresponding to high coherency (close to one). The thick black contour denotes the 5% significance level estimated from a Monte Carlo simulation of the wavelet coherency between 10,000 sets of two white noise time series with the same length studied in this paper. Within the designated areas the time series are significantly correlated at the 5% significance level. Analogously, there is a COI in cross-spectral power, representing the boundary conditions, and results within COI should be interpreted with caution. The arrows imply the phase difference between series as defined in Eq (11). Through Fig 4, we can identify both frequency bands (in the vertical axis) and time intervals (in the horizontal axis) where two indices move together. Moreover, we can explain the extent of the correlation between them across time and frequencies.

We note several findings from analyzing the co-movement patterns presented in Fig 4. First, for positive and negative sentiments, significant relationships occur primarily at medium scales and since mid-October. At higher frequencies, the significant regions are short-lived and can be due to statistical fluctuations and noise [41]. For the lead-lag relationship, it is the stock return that leads positive and negative sentiments in-phase. Fig 4 shows a significant relationship between positive sentiment and the stock market during October 20th to November 4th at the frequency of 5–10 trading days. At the beginning of October, when other peripheral markets, e.g., the United States, Europe, and Japan were undergoing frequent changes,

Chinese stock markets were closed because of the National Day holiday (October 1st to 7th). Therefore, negative sentiments were pervasive since the holiday and investors were worried about the trend of Chinese markets. During this period, fear sentiment accounted for 30% of total posts on Sina Weibo. Nevertheless, out of the widespread falling expectation, the SHCI maintained a relatively stable level, thus leading to the superficial positive effect of the stock return on negative sentiment. Hence, during this phase, the actual causation of negative sentiment is more likely to be the downtrends of peripheral markets rather than the performance of the Chinese stock market. The positive linkage between negative sentiment and the return of SHCI disappeared when China Securities Regulatory Commission announced the schedule of the Shanghai-Hong Kong Stock Connect. This landmark program provides a cross-boundary investment channel between Shanghai and Hong Kong stock markets. On October 26th, Shanghai-Hong Kong Stock Connect was postponed officially, which was interpreted as bull news by many microblogs because it implies a speculative opportunity. On the following day, the market of United States routed, followed by an increase rather than a decrease of the SCHI. The excessive rise stimulated investors' positive sentiments, and the prediction that Chinese stock market is stepping into a new adjustment which signals a bull market was pervasively reposted on Sina Weibo.

The effect of the stock market on positive sentiment sustained to November 17th, when Shanghai-Hong Kong Stock Connect was launched officially and failed to stimulate the SHCI in approximately three trading days. According to the posts on Sina Weibo, although the temporary decline of the SHCI has dampened positive sentiments inevitably, many investors insisted that Shanghai-Hong Kong Stock Connect heralds a secular bull market. With the return of SHCI increased on November 20th, positive sentiments rose significantly. The leading role of the stock market in affecting positive sentiment is strengthened by the uptrend period of late November to December, during which the daily return of SHCI reached 1.8%. However, the surge raised fear sentiments significantly with many posts expected a sharp fall to be followed. Consequently, movements in the return of SHCI since late November exert effects on both positive and negative sentiments. Whereas the positive sentiment exerts a significant instant effect on the SHCI return in June at the frequency of fewer than five trading days, no such linkage is found with negative sentiment. The result contradicts the conclusion that negative microblogs are more influential than positive microblogs in China [46] and negative posts contain more information and are more related to the stock market [26], at least from the view of stock relevant microblogs. One possible reason may be the fact that Chinese investors appear to be more overconfident than those in the United States, making them believe that past returns are indicative of future returns and increasing the trading volume and stock returns [61].

About the relationships between detailed sentiments and the stock market, we find considerable variation in relationship patterns across sentiments. Sadness shows a strong relationship with the stock market at high to medium frequencies throughout the sample period. As shown in Fig 5, significant co-movements between sadness sentiment and the return of SHCI are found during July 23th to August 20th at the frequency surrounding five trading days, and since September 5th at the frequency surrounding ten trading days. As indicated by the directions of arrows, the short-run fluctuations in sadness sentiments extracted from microblogs lead the movement of the stock market in phase during July 23th to August 20th. The medium-run fluctuations in the stock market are synchronously related to sadness with a comparatively small positive phase difference since September 5th. This result suggests that the return of SHCI is a leading indicator of sadness sentiment over the second half of the sample period. This effect may be due to the dynamic fluctuations in peripheral markets as analyzed previously. Anger presents a different pattern with less persistent co-movement

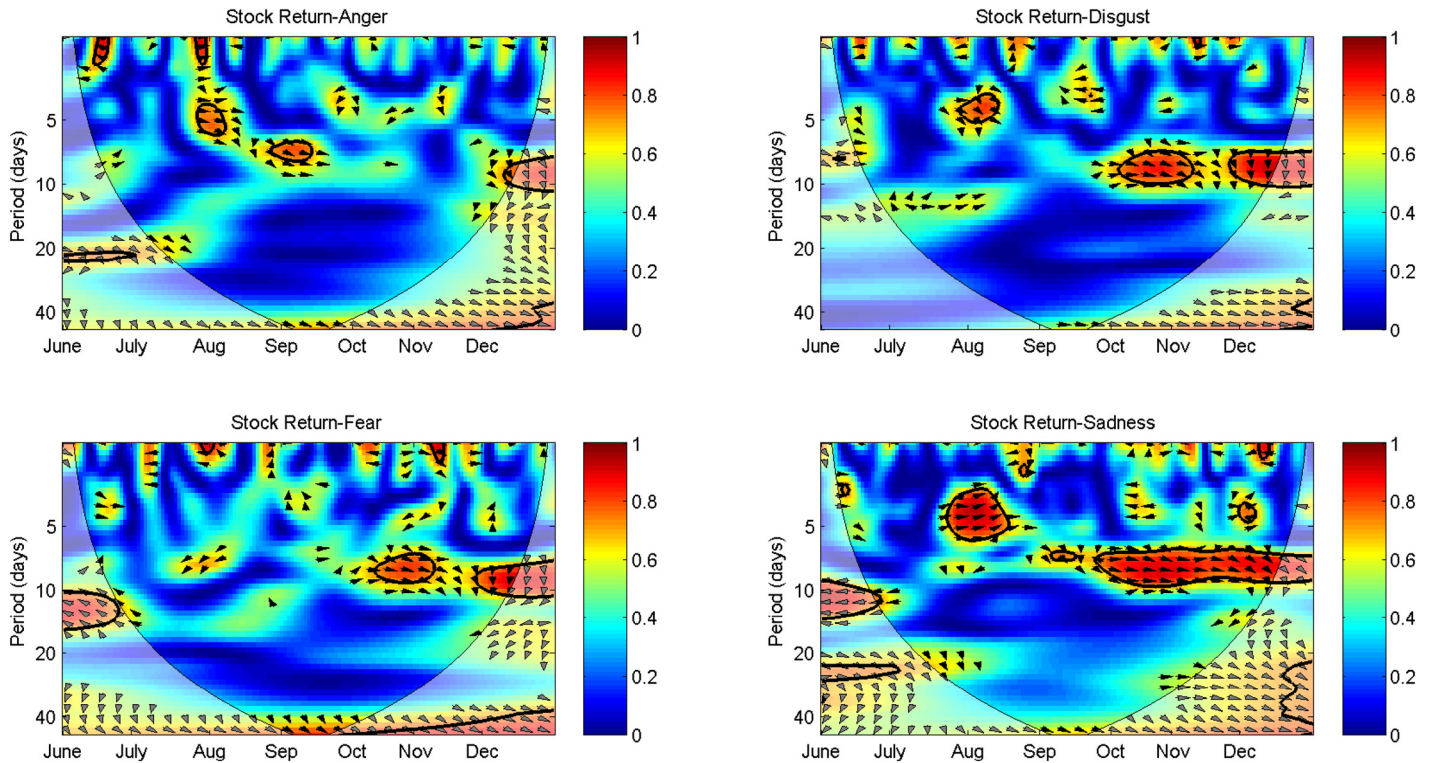


Fig 5. Morlet wavelet transform cross-spectra and coherence spectra of detailed sentiments and the SHCI return.

<https://doi.org/10.1371/journal.pone.0180723.g005>

with the stock market. During two short periods, i.e., July 28th to August 11th and August 28th to September 11th, the return of SHCI leads anger in phase. According to the mental accounting theory, faced with the sharp rise and fall in stock returns, investors are easy to segregate small gains from large losses and integrate small losses with large gains, leading to significant increase in sadness sentiments [46, 62]. Consequently, it appears that sadness exerts a closer relationship with the stock market than anger. The result conflicts with [46], in which anger sentiment is found to be more influential than sadness. Nevertheless, it is consistent with the conclusion of [36], in which sadness extracted from Sina Weibo performs better in predicting the SHCI volume than other sentiments of anger, disgust, fear, and happiness. Similarly, [37] finds that anger extracted from Sina Weibo has the weakest correlation with the Chinese stock market.

The case of fear, where the stock market acts as a leading indicator is observed mainly at the medium frequency of ten trading days, providing a similar pattern as negative sentiment, particularly during October 20th to November 14th and since November 26th. As with fear, significant relationships between disgust and the stock market appear during October 13th to November 7th and since November 26th. Affected by the fluctuating peripheral markets during Chinese National Day holiday, investors were pessimistic about the trend of SHCI. In the following 20 trading days, the peripheral markets kept fluctuating and strengthened the linkages between pessimistic sentiments and the stock market. Analogously, Shanghai-Hong Kong Stock Connect intensified the leading role of the SHCI return in affecting fear and disgust sentiments. The results are in contrast to [8] which argues that calm is the sole sentiment that causes the value of DJIA and other sentiments (alert, sure, vital, kind, and happy) have no effect on stock prices. Our analysis provides evidence that sadness has a certain leading role in

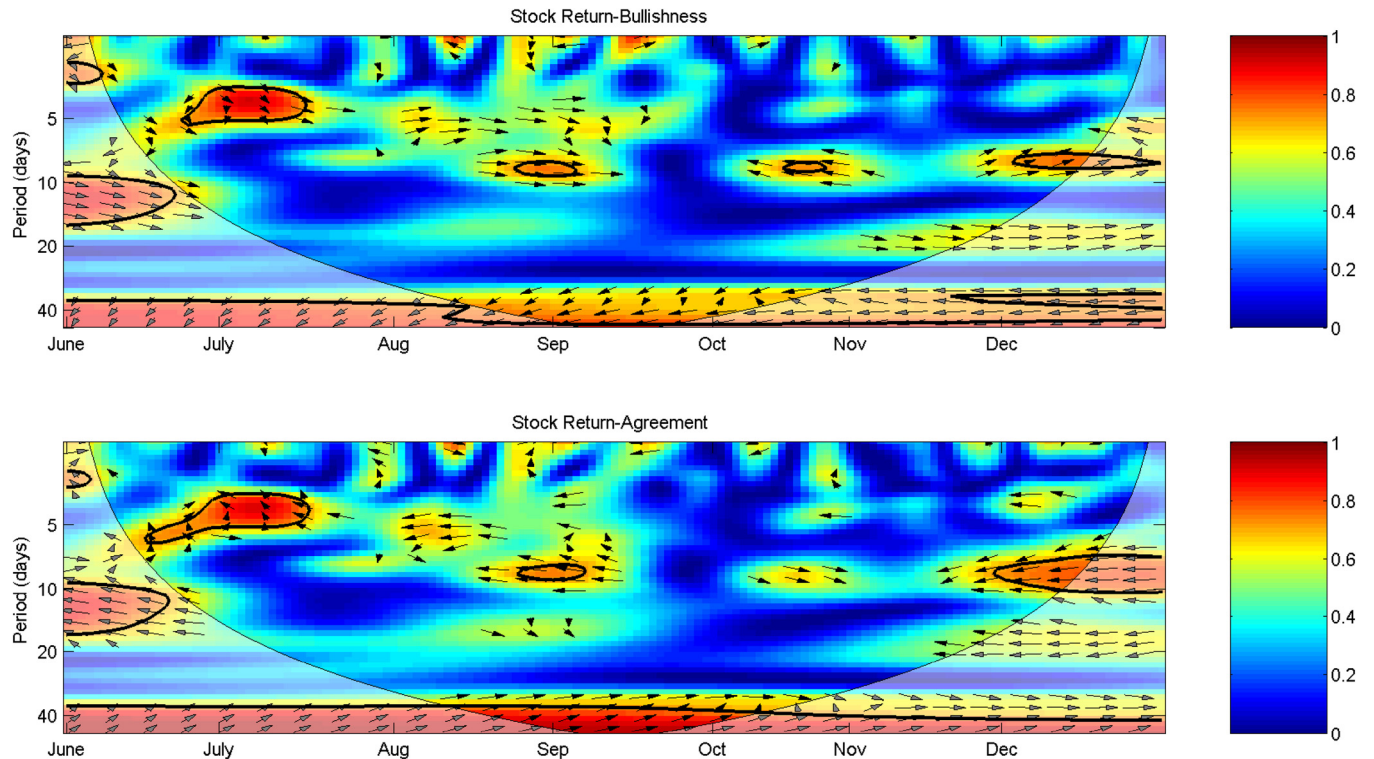


Fig 6. Morlet wavelet transform cross-spectra and coherence spectra of Bullishness/Agreement and the SHCI return.

<https://doi.org/10.1371/journal.pone.0180723.g006>

affecting the stock market, and different sentiments have distinct relationships with the return of SHCI that are time-varying and frequency related.

Individual investors are considered the least informed market participants [63]. For informed agents with limited investment capacity, they are motivated to spread informative but imprecise stock tips so that followers trade on the advice and move prices as they wanted [64]. In particular, investors are often influenced by word of mouth [65, 66, 67], reflecting their bounded rationalities in trading. Consequently, they subject to a persuasion bias, thus generating a positive relationship between the stock return and Bullishness indicator which measures the share of surplus positive signals and gives more weight to a greater number of positive posts [68]. Along with the theoretical properties, Fig 6 shows a positive phase difference between Bullishness and the SHCI return since December at the frequency surrounding ten trading days. However, no such linkage is observed over other periods. During June 25th to July 16th and August 28th to September 4th, the SHCI return leads Bullishness indicator in phase at the frequency of fewer than five trading days and ten trading days, respectively. Moreover, Bullishness leads the stock market anti-phase surrounding October 28th. Thereby, the relationship between Bullishness and the stock market is loose and unstable.

As suggested in [69], disagreement about market information leads to extensive debate and release of more information. Substantial disagreement implies that agents are uncertain about stock markets, thus causing the trading volume to rise because participants assign different values to assets [55]. Conversely, more information may not only reflect pre-existing information but also have changed market behaviors [70]. Thus, the effect of higher disagreement on the stock market is difficult to predict. Fig 6 shows the linkage between Agreement and the return of SHCI. The negative co-movements at high to medium frequencies are in line with the

results of correlation tests in [Table 1](#). During June 23th to July 17th when Agreement indicator fluctuates at the frequency of fewer than five trading days, the linkage between Agreement and the stock market is significant at the same frequency. The arrows point to the northwest, indicating that Agreement indicator leads the SHCI return anti-phase. During this period, more posts convey negative sentiments, agreeing with the indication that more agreement indicates a worse performance of the stock market. Similar results are demonstrated surrounding September. Since December, when the stock market becomes more turbulent, the linkage with Agreement changes and the SHCI return leads the indicator anti-phase. The result suggests that when the stock market is more volatile, increases in stock returns will rise the disagreement in investors. The result is in line with opinions conveyed by posts on Sina Weibo in December. The overgrowth of the SHCI at the beginning of December induced contrary expectations on future trends of the Chinese stock market. Over the long run, Agreement leads the return of SHCI in-phase at the frequency of approximately 40 trading days with arrows pointing to the northeast. The result indicates that agreements on long-run investments rise the performance of the stock market, supporting the view that less disagreement improves market certainty. A similar positive relationship between Agreement and the stock market is found by [\[57\]](#).

Overall, from the perspective of Time Domain, there are significant relationships between sentiments and the stock market in China. Nonetheless, the co-movements are unstable, and patterns of relationships are not consistent across time scales. Specifically, in most of the time during October to December when other peripheral markets are fluctuating, the SHCI return is positively correlated with Weibo sentiments, as suggested by the evidence found in [\[6\]](#) and [\[71\]](#). Over other periods, such linkage is less persistent. Whereas sadness and Agreement exert a certain leading role in affecting the stock return, the other sentiments have no such effects. From the perspective of Frequency Domain, significant relationships over low to high frequencies are demonstrated, meaning that both short-run and long-run fluctuations in sentiments and the stock market are closely related. Comparatively, the relationships of sentiments and the stock market are stronger at medium frequencies surrounding ten trading days, as indicated by the darker areas shown in the figures. Except for Agreement, other sentiments and indicators show no significant linkages at frequencies of more than 20 trading days. Moreover, results based on Wavelet analysis are in contrast to correlation tests, implying that full-sample results may be misleading and revealing that the relationships between sentiments extracted from Sina Weibo and the stock market are time-varying and frequency related.

Conclusions

This article classifies microblogs posted on Sina Weibo into five detailed sentiments (anger, disgust, fear, joy, and sadness), two polarity sentiments (positive vs. negative), and two indices representing bullishness and disagreement of stock markets (Bullishness and Agreement). Then, we assess the relationships between Weibo sentiments and the stock market in China through wavelet analysis, which is effective in capturing both frequency and time-varying features within a unified framework. There are contradictory results in the literature regarding linkages between sentiments extracted from social media and stock markets. Considering that market participants in emerging markets are more likely to be affected by psychology and sentimentality than those in developed markets, potential relationships between sentiments of investors and stock markets may be found in China. A time-frequency view of relationships between Weibo sentiments and Chinese stock market over seven months suggests that the linkage is not always significant, particularly at high frequencies. Our findings indicate that relationships between microblog sentiments and stock returns have both frequency and time-

varying features. On the one hand, the linkage between sentiments and the stock return appears to be stronger at medium-term horizons surrounding ten trading days. Compared with the other four sentiments, sadness has a closer relationship with the stock market. Polarity sentiments have certain significant relationships with the stock market but are less persistent than that of detailed sentiments. On the other hand, the agreement of investors is a leading indicator of the stock market over the long run. However, regarding the lead-lag relationship, it is the stock market that leads detailed and polarity sentiments rather than the reverse. The relationship between bullishness and the stock market is not that significant. According to the results, Chinese investors' sentiments are likely to be affected by domestic and peripheral market trends and abnormal events. Meanwhile, social media sentiments exert certain effects on Chinese market trends.

Author Contributions

Conceptualization: Yingying Xu.

Data curation: Yingying Xu, Jichang Zhao.

Formal analysis: Yingying Xu, Chiwei Su.

Investigation: Yingying Xu.

Methodology: Yingying Xu, Jichang Zhao, Chiwei Su.

Project administration: Yingying Xu, Zhixin Liu, Jichang Zhao.

Resources: Yingying Xu, Zhixin Liu, Jichang Zhao.

Software: Yingying Xu, Jichang Zhao, Chiwei Su.

Supervision: Yingying Xu, Zhixin Liu, Jichang Zhao.

Validation: Yingying Xu.

Visualization: Yingying Xu.

Writing – original draft: Yingying Xu, Zhixin Liu, Jichang Zhao, Chiwei Su.

Writing – review & editing: Yingying Xu, Jichang Zhao.

References

1. Fama EF. The behavior of stock-market prices. *J Bus.* 1965; 38(1): 34–105.
2. Fama EF. Efficient capital markets: II. *J Financ.* 1991; 46(5): 1575–1617. <https://doi.org/10.1111/j.1540-6261.1991.tb04636.x>
3. Qian B, Rasheed K. Stock market prediction with multiple classifiers. *Appl Intell.* 2007; 26(1): 25–33. <https://doi.org/10.1007/s10489-006-0001-7>
4. Nofsinger JR. Social mood and financial economics. *J Behav Financ.* 2005; 6(3): 144–160. http://dx.doi.org/10.1207/s15427579jpfm0603_4
5. Peterson RL. Affect and financial decision-making: how neuroscience can inform market participants. *J Behav Financ.* 2007; 8(2): 70–78. <http://dx.doi.org/10.1080/15427560701377448>
6. Rao T, Srivastava S. Analyzing stock market movements using Twitter sentiment analysis. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. IEEE Computer Society. 2012; August: 119–123. >10.1109/ASONAM.2012.30
7. Rechenthin M, Street WN, Srinivasan P. Stock chatter: using stock sentiment to predict price direction. *Algorithmic Finance.* 2013; 2(3–4): 169–196. <https://doi.org/10.3233/AF-13025>
8. Celov D, Grigaliuniene Ž. Economic forces, sentiment and emerging Eastern European stock markets. *Res. Econ. Bus. Central Eastern Eur.* 2013; 2(2): 37–53.
9. Lux T. Sentiment dynamics and stock returns: The case of the German stock market. *Empirical Econ.* 2011; 41(3): 663–679. <https://doi.org/10.1007/s00181-010-0397-0>

10. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. *J Comput Sci.* 2011; 2.1: 1–8. <http://doi.org/10.1016/j.jocs.2010.12.007>
11. Oliveira N, Cortez P, Areal N. On the predictability of stock market behavior using stocktwits sentiment and posting volume. *Portuguese Conference on Artificial Intelligence*; 2013 Sep 355–65; Heidelberg, Berlin. New York: Springer; 2013. https://doi.org/10.1007/978-3-642-40669-0_31
12. Si J, Mukherjee A, Liu B, Pan SJ, Li Q, Li H. Exploiting social relations and sentiment for stock prediction. In *EMNLP.* 2014; 14: 1139–1145
13. Sheu HJ, Lu YC, Wei YC. Causalities between sentiment indicators and stock market returns under different market scenarios. *Int. J. Bus. Fin. Res.* 2009; 4(1): 159–172.
14. Wang Y, Di Iorio A. The cross section of expected stock returns in the Chinese A-share market. *Global Fin. J.* 2007; 17(3): 335–349. <http://doi.org/10.1016/j.gfj.2006.05.007>
15. Zhang T, Li J, Malone P. Closed-end fund discounts in Chinese stock markets. *Chin. Econ.* 2004; 37(3): 17–38.
16. Jin X, Shen D, Zhang W. Has microblogging changed stock market behavior? Evidence from China. *Physica A.* 2016; 452: 151–156. <http://doi.org/10.1016/j.physa.2016.02.052>
17. Chen WH. Taking stock, moving forward: the Internet, social networks and civic engagement in Chinese societies. *Information, Communication & Society.* 2014; 17(1): 1–6. <https://doi.org/10.1080/1369118X.2013.857425>
18. Zhao J, Dong L, Wu J, Xu K. Moodlens: an emoticon-based sentiment analysis system for Chinese tweets. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*; 2012; Aug: 1528–1531. >10.1145/2339530.2339772
19. Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. *Seventh Conference on International Language Resources and Evaluation*; 2010; 10: 1320–1326.
20. Wysocki PD. Cheap talk on the web: The determinants of postings on stock message boards. *University of Michigan Business School Working Paper.* 1998; 98025. <http://dx.doi.org/10.2139/ssrn.160170>
21. Tumarkin R, Whitelaw RF. News or noise? Internet postings and stock prices. *Financ Anal J.* 2001; 57(3): 41–51.
22. Dewally M. Internet investment advice: Investing with a rock of salt. *Financ Anal J.* 2003; 59(4): 65–77.
23. Oh C, Sheng O. Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. *ICIS 2011 Proceedings*; 2011: 57–58.
24. Herwig U, Kaffenberger T, Baumgartner T, Jäncke L. Neural correlates of a 'pessimistic' attitude when anticipating events of unknown emotional valence. *Neuroimage.* 2007; 34(2): 848–858. <http://doi.org/10.1016/j.neuroimage.2006.09.035> PMID: [17112750](https://pubmed.ncbi.nlm.nih.gov/17112750/)
25. McDermott R, Fowler JH, Smirnov O. On the evolutionary origin of prospect theory preferences. *J Polit.* 2008; 70(2): 335–350. <https://doi.org/10.1017/S0022381608080341>
26. Garcia D, Garas A, Schweitzer F. Positive words carry less information than negative words. *EPJ Data Science.* 2012; 1(1): 3–14. <https://doi.org/10.1140/epjds3>
27. Vu TT, Chang S, Ha QT, Collier N. An experiment in integrating sentiment features for tech stock prediction in Twitter. *Proceedings of the workshop on information extraction and entity analytics on social media data. The COLING 2012 Organizing Committee*; 2012: 23–38; Mumbai, India.
28. Smailović J, Grčar M, Lavrač N, Žnidaršič M. Predictive sentiment analysis of tweets: A stock market application. *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*; 2013: 77–88; Heidelberg, Berlin; Springer. https://doi.org/10.1007/978-3-642-39146-0_8
29. Sul H, Dennis AR, Yuan LI. Trading on Twitter: The financial information content of emotion in social media. In *2014 47th Hawaii International Conference on System Sciences.* 2014; January: 806–815. <https://doi.org/10.1109/HICSS.2014.107>
30. Ranco G, Aleksovski D, Caldarelli G, Grčar M, Mozetič I. The effects of Twitter sentiment on stock price returns. *PloS one.* 2015; 10(9): e0138441. <http://dx.doi.org/10.1371/journal.pone.0138441> PMID: [26390434](https://pubmed.ncbi.nlm.nih.gov/26390434/)
31. Lin S, Ren D, Zhang W, Zhang Y, Shen D. Network interdependency between social media and stock trading activities: evidence from China. *Physica A.* 2016; 451: 305–312. <http://doi.org/10.1016/j.physa.2016.01.095>
32. Cheng WY, Lin J. Investors' bullish sentiment of social media and stock market indices. *J Manage.* 2013; 5:012.
33. Wu DD, Zheng L, Olson DL. A decision support approach for online stock forum sentiment analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems.* 2014; 44(8): 1077–1087. <https://doi.org/10.1109/TSMC.2013.2295353>

34. Xu T, Zhang H. A new approach using Weibo data to predict the China Shanghai stock market. In 2015 International Conference on Artificial Intelligence and Industrial Engineering. Atlantis Press. 2015; July.
35. Li X, Xie H, Chen L, Wang J, Deng X. News impact on stock price return via sentiment analysis. *Knowl-Based Syst.* 2014; 69: 14–23. <http://doi.org/10.1016/j.knosys.2014.04.022>
36. Dong YH, Chen H, Qian WN, Zhou AY. 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.* 2015; 7(2): 148–155. <http://dx.doi.org/10.1504/IJES.2015.069987>
37. Zhou Z, Zhao J, Xu K. Can online emotions predict the stock market in China?. In International Conference on Web Information Systems Engineering. Springer International Publishing. 2016; November: 328–342. https://doi.org/10.1007/978-3-319-48740-3_24
38. Cheng A, Evans M, Singh H. Inside Twitter: an in-depth look inside the Twitter world. Report of Sysomos. June, Toronto, Canada; 2009.
39. Guille A, Hacid H. A predictive model for the temporal dynamics of information diffusion in online social networks. Proceedings of the 21st international conference on World Wide Web; 2012; April: 1145–1152. >10.1145/2187980.2188254
40. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media?. Proceedings of the 19th international conference on World Wide Web; 2010; April: 591–600. >10.1145/1772690.1772751
41. Kristoufek L. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PLoS one.* 2015; 10(4): e0123923. <http://dx.doi.org/10.1371/journal.pone.0123923> PMID: 25874694
42. Dodds PS, Danforth CM. Measuring the happiness of large-scale written expression: songs, blogs, and presidents. *J Happiness Stud.* 2010; 11(4): 441–456. <https://doi.org/10.1007/s10902-009-9150-9>
43. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS one.* 2011; 6(12): e26752. <http://dx.doi.org/10.1371/journal.pone.0026752> PMID: 22163266
44. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science.* 2011; 333(6051): 1878–1881. <https://doi.org/10.1126/science.1202775> PMID: 21960633
45. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. CS224N Project Report. Stanford; 2009.
46. Fan R, Zhao J, Chen Y, Xu K. Anger is more influential than joy: Sentiment correlation in Weibo. *PloS one.* 2014; 9(10): e110184. <http://dx.doi.org/10.1371/journal.pone.0110184> PMID: 25333778
47. Liu B. Sentiment analysis and subjectivity. *Handbook of natural language processing.* 2010; 2: 627–666.
48. Grinsted A, Moore JC, Jevrejeva S. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Proc Geoph.* 2004; 11: 561–566.
49. Rua A. Money growth and inflation in the Euro Area: A time-frequency view. *Oxford B Econ Stat.* 2012 (74): 875–885. <https://doi.org/10.1111/j.1468-0084.2011.00680.x>
50. Wang J, Zhu J, Dou F. Who plays the key role among Shanghai, Shenzhen and Hong Kong stock markets? *China World Econ.* 2012; 20: 102–120. <https://doi.org/10.1111/j.1749-124X.2012.12004.x>
51. Hudgins L, Friehe CA, Mayer ME. Wavelet transforms and atmospheric turbulence. *Phys Rev Lett.* 1993; 71: 32–79. <https://doi.org/10.1103/PhysRevLett.71.3279>
52. Torrence C, Webster PJ. Interdecadal changes in the ENSO-monsoon system. *J Climate.* 1999; 12: 2679–2690. [http://dx.doi.org/10.1175/1520-0442\(1999\)012<2679:ICITEM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1999)012<2679:ICITEM>2.0.CO;2)
53. Bloomfield DS, McAteer RJ, Lites BW, Judge PG, Mathioudakis M, Keenan FP. Wavelet phase coherence analysis: application to a quiet-sun magnetic element. *Astrophys J.* 2004; 617(1): 623–632. <https://doi.org/10.1086/425300>
54. Antweiler W, Frank MZ. Is all that talk just noise? The information content of internet stock message boards. *J Financ.* 2004; 59(3): 1259–1294. <https://doi.org/10.1111/j.1540-6261.2004.00662.x>
55. Harris M, Raviv A. Differences of opinion make a horse race. *Rev Financ Stud.* 1993; 6(3): 473–506. <https://doi.org/10.1093/rfs/5.3.473>
56. Milgrom P, Stokey N. Information, trade and common knowledge. *J Econ Theory.* 1982; 26(1): 17–27. URL: [https://doi.org/10.1016/0022-0531\(82\)90046-1](https://doi.org/10.1016/0022-0531(82)90046-1)
57. Sprenger TO, Tumasjan A, Sandner PG, Welpe IM. Tweets and trades: The information content of stock microblogs. *European Financial Management.* 2014; 20(5): 926–957. <https://doi.org/10.1111/j.1468-036X.2013.12007.x>
58. Brown GW, Cliff MT. Investor sentiment and the near term stock market. *J Empir Financ.* 2004; 11: 1–27. <http://doi.org/10.1016/j.jempfin.2002.12.001>

59. Brown GW, Cliff MT. Investor sentiment and asset valuation. *J Bus.* 2005; 78(2): 405–439. <https://doi.org/10.1086/427633>
60. Baker M, Wurgler J. Investor sentiment and the cross-section of stock returns. *J Financ.* 2006; 61: 1645–1680. <https://doi.org/10.1111/j.1540-6261.2006.00885.x>
61. Chen G, Kim KA, Nofsinger JR, Rui OM. Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. *J Behav Decis Making.* 2007; 20(4): 425–451. <https://doi.org/10.1002/bdm.561>
62. Gou Y, Jiang Y, Rui L, Miao D, Peng J. The nonfungibility of mental accounting: A revision. *Soc Behav Personal.* 2013; 41(4): 625–633. <https://doi.org/10.2224/sbp.2013.41.4.625>
63. Hirshleifer D, Teoh SH. Limited attention, information disclosure, and financial reporting. *J Account Econ.* 2003; 36(1): 337–386. <http://doi.org/10.1016/j.jacceco.2003.10.002>
64. Bommel JV. Rumors. *J Financ.* 2003; 58(4): 1499–1520. <https://doi.org/10.1111/1540-6261.00575>
65. Hong H, Kubik JD, Stein JC. Thy neighbor's portfolio: word-of-mouth effects in the holdings and trades of money managers. *J Financ.* 2005; 60(6): 2801–2824. <https://doi.org/10.1111/j.1540-6261.2005.00817.x>
66. Ng L, Wu F. Peer effects in the trading decisions of individual investors. *Financ Manage.* 2010; 39(2): 807–831. <https://doi.org/10.1111/j.1755-053X.2010.01093.x>
67. Mizrach B, Weerts S. Experts online: an analysis of trading activity in a public Internet chat room. *J Econ Behav Organ.* 2009; 70(1): 266–281. <http://doi.org/10.1016/j.jebo.2009.02.001>
68. DeMarzo PM, Vayanos D, Zwiebel J. Persuasion bias, social influence, and unidimensional opinions. *Q J Econ.* 2003; 118(3): 909–968. <https://doi.org/10.1162/00335530360698469>
69. Das SR, Martinez-Jerez A, Tufano P. E-Information: A clinical study of investor discussion and sentiment. *Financ Manage.* 2005; 34(3): 103–137. <https://doi.org/10.1111/j.1755-053X.2005.tb00112.x>
70. Jones AL. Have internet message boards changed market behavior?. *Info-The journal of policy, regulation and strategy for telecommunications.* 2006; 8(5): 67–76. <https://doi.org/10.1108/14636690610688105>
71. Bouktif S, Awad MA. Ant colony based approach to predict stock market movement from mood collected on Twitter. In *Advances in Social Networks Analysis and Mining (ASONAM)*. 2013; 837–845. <https://doi.org/10.1145/2492517.2500282>