Quantifying the Effects of Online Bullishness on International Financial Markets

Huina Mao[1], Scott Counts[2], and Johan Bollen[1]

[1]School of Informatics and Computing, Indiana University, Bloomington, Indiana, United States
[2]Microsoft Research, Redmond, Washington, United States

Abstract

Computational methods to gauge investor sentiment from large-scale online data sources using machine learning classifiers and lexicons have shown considerable promise, but suffer from measurement and classification errors. In our work we develop a simple, direct, and unambiguous indicator of online investor sentiment, which is extracted from Twitter updates and Google search queries. We examine the predictive power of this new investor Bullishness indicator on international stock markets. Our results indicate several striking regularities. First, changes in Twitter bullishness predict changes in Google bullishness, indicating that Twitter information precedes Google queries. Second, Twitter and Google bullishness are positively correlated with and *lead* the investor sentiment survey. Especially, the former has greater stock market predictive value than the latter. Third, we observe high Twitter bullishness predicts increases of stock returns, followed by a reversal to the fundamentals. We speculate that our results support the investor sentiment hypothesis in behavioral finance.

Introduction

The Efficient Market Hypothesis (EMH) [1] states that investors operate as rational actors, and that stock market prices therefore fully reflect *all* existing, new, and even hidden information. However, traditional efficient-market models fail to explain important market anomalies, such as the Great Crash of 1929, the Black Monday Crash of October 1987, the late 90s Dot-com bubble, and the market collapse of 2008. Behavioral finance challenges the EMH by emphasizing the important role of behavioral and emotional factors in investor behavior [2, 3]. Behavioral finance has two major assumptions, namely "investor sentiment", i.e. investors are subject to sentiment, not just rational considerations, and "limits-to-arbitrage", i.e. betting against irrational investors is costly and risky. Due to the limited arbitrage of sophisticated investors, investor sentiment can influence stock prices [4]. The quantification and measurement of investor sentiment, and its effects, has therefore become an important research topic [5].

In recent years, researchers have explored a variety of computational methods to measure large-scale market sentiment indicators from online data sources, such as investor message boards, news, micro-blogging environments, blogs, and search engine query streams. This approach holds considerable promise, given the underlying data's unprecedented large scale, high resolution, low cost, and high frequency.

To the best of our knowledge, existing market sentiment measures can be categorized into two main classes, namely classifier- and dictionary-based. In [6], two popular classifiers – Naive Bayes and Support Vector Machine – are employed to classify stock messages into three categories, namely *bullish*, *bearish*, and *neutral*. Their research has found that message bullishness and volume help predict market volatility, but has modest value to predict returns. Similar results have been obtained in later work that uses as many as five classifier algorithms [7]. The latest and most relevant study [8] classifies stock tweets from Stocktwits.com into *bullish* and *bearish* categories, and builds a bullishness index that is shown to be predictive of the future stock price movement.

Along with machine-learning approaches, a number of approaches have focused on the development of linguistic lexicons, or dictionaries, to determine investor sentiment from word frequencies in financial data sources. Perhaps the most influential study is Tetlock's [9] who determine the frequency of words in the Harvard Negative word list in daily news to construct a pessimism indicator which was found to predict the daily Dow Jones returns and the firm stock prices reported in his following work [10]. However, the authors in [11] argue that the Harvard psychosocial dictionary is developed for the domains of psychology and sociology, hence many words that are classified as negative are not negative in a financial context. They developed an alternative negative word list that contained 2,337 words and was shown to outperform the Harvard dictionary in measuring financial sentiment.

Classifier- and dictionary-based methods are useful to *automatically* process large-scale text data for the extraction of general sentiment indicators. However, the variegated contexts and subtleties of human language pose a tough challenge to human raters and text analysis algorithms. In fact, the low accuracy with which *humans* themselves can assess text sentiment inevitably sets an unfavorable upper bound on what the best supervised classifiers can achieve. According to [7,8,12], a machine learning classification accuracy of 60 - 70% is considered to be acceptable. Dictionary-based methods do not require human-defined ground truth or supervision, but dictionary words are usually selected on the basis of ad hoc criteria, the word weighting schemes may be biased and context-sensitive, and dictionaries can not be adjusted to varying word context and semantics.

The limitations of automated sentiment analysis algorithms are not merely an academic or technical matter. Investors are averse of ambiguity and uncertainty [13]. For computational indicators of investor sentiment to become an accepted part of the financial tool kit, they need to be reliable, accurate, and reduce ambiguity and risk rather than to increase it.

Compared to computational indicators, surveys of investor sentiment have already established themselves as an accepted part of the financial data ecology. For example, Daily Sentiment Index (DSI) and weekly Investor Intelligence (II) are two well-known surveys of investor sentiment. Since 1987, DSI interviews small traders for their bullish or bearish feelings on US future markets. Since 1963, II asks and categorizes readers' opinion from market newsletters into three categories, namely bullish, bearish, or correction, i.e. neutral. Simply put, surveys measure bullish or bearish sentiment from what people explicitly *tell* others when asked. This certainly has the advantage of being unambiguous and precise, but surveys can still be subject to a number of detrimental shortcomings: they are resource intensive and expensive to conduct. Furthermore, what may seem a strength could actually be a weakness; when explicitly asked for their opinion variety of individual and social biases, including group-think, and respondents' truthfulness can become an issue [14, 15].

Here we aim to define an indicator of investor sentiment that maintains the advantage of traditional surveys by requiring explicit, unambiguous statements of investor sentiment, yet leverages large-scale online social media data. Our indicator measures investor sentiment directly from what people *tweet* or *search*, rather than what they *tell* others in response to survey questions. To reduce the ambiguities of sentiment analysis, we measure the relative occurrences of only two terms: *"bullish"* and *"bearish"* which were chosen because they are rarely used other than in financial contexts. They are thus more likely to produce an unambiguous indication of bullish or bearish investor sentiment.

In this paper we collect the frequencies of the terms "bullish" vs. "bearish" from Twitter content and Google queries over time, and define a bullishness index on the basis of their relative frequencies. We compare the Bullishness indicators calculated respectively from Twitter content and Google queries, i.e. same index, but different data sources. We also compare both with existing surveys of investor sentiment, and examine their predictive effect on the stock market returns across the United States (US), the United Kingdom (UK), Canada (CA), as well as China (CN). Our results indicate a positive correlation between survey sentiment and Twitter & Google Bullishness. Twitter bullishness has statistically and economically significant predictive value towards US, UK and CA market prices. We further observe that high Twitter bullishness predicts the increase of daily returns on the next day followed by a reversion in the next 2-5 days. Our results support the investor sentiment theory [4], and suggest that Twitter bullishness may be a useful and simple investor sentiment index.

Results

Twitter Bullishness

We define a tweet as Bullish if it contains the term "bullish" and Bearish if it contains the term "bearish". Over the study period of 2010 to 2012, we find about 0.31 million bullish and bearish tweets. There are 1,091 days in total, and the daily average number of bullish and bearish tweets is 280. Fig. 1 shows bullish & bearish tweet volume. The autocorrelation graph in the left panel indicates a clear weekly pattern, which is also confirmed by the Fast Fourier Transform result shown in the right panel. In the magnitude spectrum plot, the first dominant peak indicates the whole period as the main periodicity, while the second and third ones appear at 6.99 days and 3.50 days respectively. So, the time series of bullish and bearish tweet volume exhibits a strong weekly pattern, with high volumes during trading days (weekdays), a peak on Tuesday and Thursday, and lower volumes during non-trading days (weekends). This finding is consistent with earlier studies [8] that suggests the distribution of bullish or bearish messages matches investor behavior. The average ratio of the number of bullish tweets over the sum of number of bullish and bearish tweets is 69.4%, suggesting either a bias toward optimism on the part of online investors [8] or an effect of the Pollyanna Hypothesis [16] which posits that humans universally favor positive words over negative words.

Following earlier work [6,8], we define a Twitter Bullishness index whose value on day t is given by Eq. 1.

$$B_t = \ln\left(\frac{1+||\mathcal{B}_t||}{1+||\mathcal{R}_t||}\right) \qquad G_w = \ln\left(\frac{1+||\mathcal{B}_w||}{1+||\mathcal{R}_w||}\right) \tag{1}$$

$\mathcal{B}_t$ and $\mathcal{R}_t$ denote the sets of bullish and bearish tweets on day t, respectively. The logarithmic transformation attenuates the effect of extremely large numbers of tweets. Studies have shown that this particular form outperforms two alternatives [6].

Google Bullishness

In a similar fashion to Twitter Bullishness B_t, we define Google Bullishness G_w in Eq. 1 from the volumes of Google queries that contain the corresponding financial terms. The volume of such queries is retrieved from Google Trends, necessitating a few notable changes. First, we found that Google search volumes of the adjectives "bullish" and "bearish" are insignificant, likely because isolated adjectives are rarely searched for by Google users. Google's Hot Trends indeed indicates that the overwhelming majority of search queries are nouns. We therefore chose to replace the adjectives "bullish" and "bearish" with their equivalents "bull market" and "bear market" for our Google Bullishness G_w indicator. The latter provide better coverage (see Fig. 2). For China we record the Mandarin ideograms "牛市" (i.e. bull market) and "熊市" (i.e. bear market). Second, Google search volumes are only available on a weekly basis whereas Twitter volume can be recorded at any temporal resolution. Google Bullishness G_w on the week w is therefore defined in Eq. 1 as the weekly ratio of $||\mathcal{B}_w||$ and $||\mathcal{R}_w||$ which represents the search volumes of "bull market" and "bear market" on the w^{th} week, respectively.

International Stock Markets

In this paper we compare Twitter and Google's Bullishness to stock market values across four different countries to increase the robustness of our results, namely the United States (US), the United Kingdom (UK), Canada (CA), and China (CN) which were selected for the following reasons. First, they are large market capitalization countries in the world, according to the World Bank statistics reported in 2012 (`http://data.worldbank.org/indicator/CM.MKT.LCAP.CD`). Second, both Google and Twitter enjoy widespread adoption in the US, UK, and CA. Therefore, online behavior as measured from Twitter and Google in these countries are more likely to be representative of trends in the general population. Third, we deliberately included China in our study because its investor behavior, market structure, legal system as well as the uptake of social media and search engines is quite different in China from the US, UK and Canada. It can therefore increase the diversity and robustness of our study.

We represent each nation's stock market by a selected index, i.e. the Dow Jones Industrial Average (DJIA) for the US, the FTSE 100 Index for UK, the S&P/TSX Composite Index (GSPTSE) for Canada, and the SSE Composite Index for China. The monthly stock prices of these four countries are shown in Fig. 3.

Research Questions and Inference

We specifically set out to address three research questions. First, are Twitter and Google Bullishness related? Although one is derived from daily micro-blogging updates on Twitter and the other from weekly Google search query volumes, both originate from online activity and may as a result reflect similar features of online investor sentiment. Second, since Twitter is a rather fast-response, online medium, indicative of rapid changes in news and sentiment, does Twitter Bullishness lead or lag daily stock market returns? Third, since the same applies to Google query streams, does Google Bullishness lead or lag weekly stock market returns? Throughout, we will control survey-based measurements of investor sentiment in our prediction analysis.

Lead-Lag Relation between Twitter and Google Bullishness

We compare Twitter Bullishness and Google Bullishness over time, and determine whether they are correlated, and whether one leads or lags other.

As shown in Eq. 1, Google Bullishness G_w is a weekly time series vs. Twitter Bullishness B_t which is a daily time series; Google query data is only available weekly from Google Trends, whereas Twitter data can be collected at any time interval. In order to compare G_w and B_t at the same time scale, we calculate the weekly mean of Twitter Bullishness, denoted B_w. The sample period thus includes 156 weeks from January 9th, 2010 to December 29th, 2012.

We find a positive and statistically significant correlation between Twitter and Google Bullishness ($\gamma = +0.27, p = 0.0007$). To estimate the lead-lag relation between the two bullishness indexes in both directions we use a Vector Autoregression (VAR) framework. VAR is a linear statistical model that captures the inter-dependencies among multivariate time series, and is widely used to validate and quantify the predictability of financial indicators [9, 17, 18]. Our VAR model is equivalent to the bivariate Granger Causality test proposed in [19], and is shown in Eq. 2.

$$\Delta G_w = \alpha + \sum_{i}^{4} \beta_i \Delta G_{w-i} + \sum_{i}^{4} \chi_i T_{w-i} + \epsilon_w \tag{2}$$

The historical lag is chosen to be 4 weeks. Since VAR is sensitive to non-stationarity, we conduct an augmented Dickey-Fuller Test which indicates that G_w is non-stationary, while B_w is stationary at a 90% confidence level. Therefore, we take the first order difference of Google Bullishness which we denote ΔG_w.

All variables in our regression model are normalized to standardized scores. Table 1 lists coefficient estimates with p-values. The reported coefficients measure the impact of one standard deviation increase of an independent variable on the change of Google Bullishness in the week w. ϵ_t is found to satisfy the linear regression assumptions: independence, homoscedasticity, and normality.

From Table 1, we can see Twitter Bullishness has a statistically significant and positive influence on the change of Google Bullishness in the following week. But ΔG_{w-1} and ΔG_{w-2} are negatively related to the change of Google Bullishness ΔG_w. We speculate that the negative sign may be the result of limitations in human attention spans [20], i.e. Google users may switch their search attention from one topic to another in the span of 2 or 3 weeks.

We note that only 23% of the variance of ΔG_w can be explained indicating the difficulty of prediction from out of sample data sources like Twitter and Google. In addition, when we reverse the regression direction, we do not find any significant prediction relation from G_w to B_w, i.e. Twitter Bullishness *leads* Google Bullishnes, but not vice versa. This finding may indicate a potential efficiency gain of Twitter over Google search, but we leave it to future research to examine and potentially explain the latter effect in more detail.

Twitter Bullishness vs. Stock Market Returns

Given that B_t leads G_w we first apply the VAR model to examine whether Twitter Bullishness has predictive value with respect to stock market returns.

First, we study the US stock market which is the largest in the world. Furthermore, the US has the highest concentration of Twitter users in the world. There are several major US market indexes, including Dow Jones Industrial Average (DJIA), Standard & Poor's 500 (SP500) and Russell 3000. DJIA, SP500, and Russell 3000 contain the 30, 500 and 3,000 largest companies, respectively. Russell 3000 Index can be further divided into large-cap Russell 1000, i.e. the top 1,000 companies, and small-cap Russell 2000, i.e. the bottom 2,000 companies. To test the robustness of our method, we examine Twitter Bullishness prediction on all the major US stock indexes.

The log stock return (R_t) is calculated on the basis Eq. 3.

$$R_t = \log(S_t^{close}) - \log(S_t^{open}) \tag{3}$$

where S_t^{close} and S_t^{open} are the stock market closing and open prices on day t, respectively. Since the daily Twitter Bullishness is calculated from 00:00 to 23:59:59 Greenwich Mean Time (GMT), while daily US market returns are computed from 16:00 to 15:59:59 Eastern Time (ET), the log return R_t is calculated from open price to closing prices on day t to avoid the possibility of including after-hour information that may not be fully reflected in the next day's closing price.

To evaluate the contribution of any new predictor such as Twitter Bullishness we need to control for existing predictors. In line with earlier work [9], the endogenous variables of our model include the stock price as well as trading volume to take into account liquidity effects. Log trading volume is de-trended to ensure stationary. The third endogenous variable is our Twitter Bullishness index B_t. The exogenous variables include VIX (the "fear index"), Daily Sentiment Index (DSI); a proxy for investor sentiment, and calendar controls, including dummy variables for Monday and January. All variables in the model are lagged up to five days which corresponds to one trading week.

The regression model is thus defined as:

$$R_t = \alpha + \sum_i^5 \beta_i R_{t-i} + \sum_i^5 \chi_i T_{t-i}^B + \sum_i^5 \delta_i Vol_{t-i} + \phi_i Exog_t + \epsilon_t \tag{4}$$

Table 2 shows the regression coefficient estimates and associated p-values. Each coefficient indicates the impact of one standard deviation increase in Twitter Bullishness on daily returns in basis points (1 basis point equals 0.01% of a daily return). The Durbin-Watson statistic for the regression residual (ϵ_t) is DW $= 2, p = 0.5$, indicating near absence of autocorrelation. In addition, ϵ_t in the model is found to be normally distributed.

The first column of Table 2 lists the regression estimation for Dow Jones. We observe that one standard deviation increase of Twitter Bullishness on day $t-1$ is followed by 12.56 basis points (bps) increase in DJIA returns on the next day. This impact is statistically significant at the 99% confidence level. In addition, comparing with the unconditional mean of daily Dow Jones returns during the sample period that is 3.46 bps, 12.56 bps is also economically significant. We also compare Twitter Bullishness to a survey of investor sentiment, i.e. Daily Sentiment Index, for their contemporaneous correlations and predictive effect on stock returns. The Pearson correlation coefficient between DSI and Twitter Bullishness ($\gamma = 0.30$, $p \ll 0.01$) is statistically significant but not high. We also found that one-standard-deviation increase in DSI is followed by only 2.26 bps increase of daily Dow returns, which is not economically significant and only marginally statistically significant with $p = 0.1, t = 1.6$. This result suggests that Twitter Bullishness, as a new proxy for investor sentiment, is related to but different from existing DSI, and can have larger predictive effects on the stock stock market than survey-based indicators.

To examine the robustness of the Twitter Bullishness' predictive value we performed further tests vs. the large-cap SP500, large-cap Russell 1000, and small-cap Russell 2000. The results of this analysis are reported in the 2^{nd}-4^{th} columns of Table 2, respectively. It is found that Twitter Bullishness of the previous day has statistically and economically significant effects on SP500, Russell 1000, and Russell 2000. Moreover, we observe a price reversal on the 4th day lag for these four market indexes, even though it is not statistically significant for DJIA and SP500. In particular, for Russell 1000 and Russell 2000, the initial increases on the first day are almost completely offset by the reversal in the lag 4. Our finding is consistent with the investor sentiment model [4], which claims that noise traders' irrationality can drive the asset price to deviate from its fundamental value temporarily after which it will reverse to the mean.

Besides the US stock market, we test the predictive value of Twitter Bullishness on the stock markets of the United Kingdom (UK), Canada (CA) and China (CN). Twitter enjoys widespread adoption in the UK and Canada, so one may expect that Twitter Bullishness may contain relevant information for the UK and CA stock markets as well. Unlike UK and Canada, Twitter is not used in China. The comparison between Twitter Bullishness and the Chinese stock market can therefore serve as a null-model, i.e. one would expect that Twitter Bullishness has much less forecasting power for the Chinese stock market than other countries. We use the VAR model to validate our assumptions.

Due to limited availability of existing predictive indicators for UK, CA and CN markets, we adopt a reduced regression model in Eq. 5 to examine the forecasting power of Twitter Bullishness vs. the stock markets of these countries.

$$R_t = \alpha + \sum_{i}^{5} \beta_i R_{t-i} + \sum_{i}^{5} \chi_i T_{t-i}^{B} + \epsilon_t \tag{5}$$

Daily returns are computed based on the main stock market index of these countries, namely DJIA for US, FTSE100 for UK, GSPTSE for CA, and SSE for CN. The regression coefficient estimates are reported in Table 3. The coefficient measures the impact of one standard deviation increase of Twitter Bullishness on daily returns in basis points.

We find that both the reduced model in Eq. 5 and the full model of Eq. 4 generate nearly the same results in terms of the Twitter Bullishness predictive value vs. the DJIA. The impact of one standard deviation of Twitter Bullishness on next day Dow Jones is about 13 basis points in both models. Adding controls into the full model does not seem to harm the predictability of Twitter Bullishness, which again indicates that Twitter Bullishness may contain relevant information for market prediction that is not captured by existing variables.

Further indicating is that our results are robust, we find similar predictive value of Twitter Bullishness vs. the UK and CA stock markets. We observe similar reversal effects that are furthermore stronger for the UK and CA than the US. With respect to predicting China's financial markets, we find that Twitter Bullishness has a much lower predictive value (8.73 bps) with only marginal statistical significance ($p = 0.09$). This may be because Twitter is closed down in China. Instead, Weibo is the most popular microblogging platform in China.

Google Bullishness vs. Stock Market Returns

We obtain the search volumes of "bull market" and "bear market" from Google Trends from January 2007 to December 2012, which constitutes 313 data points (weeks) in total. Google Bullishness is calculated based on Eq. 1. Fig. 4 plots the trend of the stock market index prices against Google Bullishness. We track the search volumes of "bull market" and "bear market" both in English and

Chinese. Chinese Google Bullishness is constructed based on the search volume of the ideograms "牛市" (i.e. bull market) and "熊市" (i.e. bear market).

The Pearson linear correlation coefficients between Google Bullishness and the corresponding log stock market prices of US, UK, CA and CN are 0.30, 0.38, 0.23 and 0.65, which are all statistically significant ($p \ll 0.01$). From Fig. 4, one can observe the positive relation between Google Bullishness and stock price levels. In addition, the former seems to lead the latter. Interestingly, this is particularly the case at market extremes. For example, Google Bullishness touched a bottom in middle 2008 before a market crash in late 2008 and early 2009 in US, UK and CA. In a similar fashion, Chinese Google Bullishness reached a peak in early 2007 that preceded a market peak in early 2008. Subsequently, a declining trend of Bullishness is followed by a down trend of the market until 2009. It is surprising to find that Chinese Google Bullishness has the highest correlation ($\gamma = 0.65$) with Chinese market relative to the markets of the other three countries under consideration where Google is the leading search engine. In China, Google only has about less than 15% search market share in 2012 compared to Baidu that owns over 75%. The stronger positive correlation between Chinese stock market and Google Bullishness may be attributed to the large population of Chinese Internet users (in 2012 there are over 500 million Internet users in China). This result is highly suggestive of the potential to study the value of online sources for Chinese market prediction, a topic that has received less interest in the literature.

Significant correlations between Google Bullishness and stock prices do not tell us whether one leads the other. Following the same regression framework adopted above, we investigate the predictability of weekly Google Bullishness on market returns, i.e. the difference between the log closing price of this week and last week. However, both the level and the change of Google Bullishness are not predictive of the weekly returns of US.DJIA, UK.FTSE100, CA.GSPTSE, and CN.SSE (see Table 5).

We note that the lack of predictive value of Google Bullishness vs. the financial markets under investigation, may be explained by the fact that Google Trend data is provided at a weekly time scale. Over that time span the market is likely to incorporate useful information and adjust prices accordingly, therefore Google Bullishness being derived from weekly Google Trend data would not contain predictive information.

In the reverse direction, we test the impact of weekly returns on the level and the change of Google Bullishness. The results are highly statistically significant. This finding supports the positive feedback trading theory in [21], i.e. traders' optimism increases when stock prices increase, and vice versa traders' pessimism increases when the prices decrease.

Despite the failure in predicting weekly stock returns, we test whether Google Bullishness may convey predictive information of *investor sentiment* rather than market prices. Investor Intelligence (II) is a well-accepted investor sentiment index in finance that measures whether US financial advisors' sentiment is bullish, bearish, or neutral. Based on Eq. 1, we compare II Bullishness to our Google Bullishness. Fig. 5 displays the trend of Google and II Bullishness and their cross-correlation results. For the lags in the range of [-3 to 3], the correlation corefficients are 0.34, 0.40, 0.47, 0.54, 0.59, 0.60 and 0.59, correspondingly.

The linear correlation between II Bullishness and Google Bullishness measured from United States is highly positive: $\gamma = 0.54$, $p \ll 0.01$. More importantly, from the cross correlation results in Fig. 5, we observe that Google Bullishness may in fact lead II Bullishness. We use VAR to estimate the predictive relation between these two different sentiment indicators. The time series are de-trended to be stationary by taking first order difference. The result is shown in Table 4.

The residuals in this model have no significant autocorrelation (Durbin-Watson statistic = 2.0; $p = 0.5$), and meet the other two model assumptions of homogeneity and normality. Surprisingly, the lagged values of II Bullishness do not carry any predictive power by themselves, whereas Google Bullishness does in lags ranging from 1 to 3 weeks. However, the regression model only explains about 6% of the variance, indicating difficulty in predicting change of investor sentiment from these variables.

Discussion

The reliability and accuracy of existing computational measures of investor sentiment leaves much to be desired. We therefore propose a direct and unambiguous measure of investor sentiment, namely the relative frequency of occurrence of two commonly terms used by investors terms in Twitter updates and Google queries. Daily Twitter Bullishness is indeed found to be an useful investor sentiment indicator. Our analysis shows a positive correlation between Twitter Bullishness and Google Bullishness on a weekly basis, and finds furthermore that the former leads changes in the latter. In addition, the two indicators of Bullishness from different data sources are found to be positively correlated with existing surveys of investor sentiment, such as Daily Sentiment Index and Investor Intelligence. More importantly, we find that daily Twitter Bullishness leads the US stock index returns (Dow Jones, SP500, Russell 1000, and Russell 2000), as well as the UK FTSE 100, and Canada GPSTSE, while having only very modest predictive value with respect to the Chinese stock market, as expected. Although high Twitter Bullishness predicts the increase of stock returns, we do observe a reversion to fundamental values during the first week. Our research thus seems to support the hypothesized role of "investor sentiment" in behavioral finance. We also note the strong positive linear correlation between Google Bullishness and Chinese stock prices ($\gamma = 0.65, p \ll 0.01$), where the former seems to lead the latter at market extremes. This result is highly suggestive of the potential to study the value of online sources such as Weibo for Chinese market prediction, a topic that has received less interest in the literature.

Methods

Twitter and Google Bullishness

We derive Twitter and Google Bullishness scores based on the volume of bullish and bearish tweets and search queries. We simply select words "*bullish* or *bull market*" and "*bearish* or *bear market*" to identify bullish and bearish sentiment, because they are rarely used in non-financial contexts, and their meanings are relatively unambiguous. The definition of the online Bullishness index is shown in Eq. 1.

Data retrieval

Our Twitter dataset is mainly acquired via Twitter Gardenhose, which consists of a random sample of public tweets (about 45 million tweets per day) during the time period of January 2010 to December 2012. Google search query data is retrieved from Google Trends (`http://www.google.com/trends/`) in 2012, which provides weekly search volume data from January 2004 to the present for any given query. Values are dynamically scaled to the range of [0, 100], between volume peaks and troughs. Two investor sentiment surveys, Daily Sentiment Index (DSI) (`http://www.trade-futures.com/dailyindex.php`) and Investor Intelligence (`http://www.investorsintelligence.com/x/us_advisors_sentiment.html`), were kindly made available to our investigation. All the historical market data is retrieved from Yahoo Finance! (`http://finance.yahoo.com/`) in 2012.

References

[1] Malkiel, B. G. & Fama, E. F. Efficient capital markets: A review of theory and emiprical work. *The Journal of Finance* **25**, 383–417 (1970).

[2] Kahneman, D. & Tversky, A. Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263–291 (1979).

[3] Shiller, R. J. *Irrational Exuberance* (Crown Business, 2006).

[4] Long, B., Shleifer, A., Summers, L. & Waldmann, R. Noise trader risk in financial markets. *Journal of Political Economy* **98**, 703–738 (1990).

[5] Baker, M. & Wurgler, J. Investor sentiment in the stock market. *Journal of Economic Perspectives* **21**, 129–152 (2007).

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

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

[8] Oh, C. & Sheng, O. R. L. Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. *ICIS 2011 Proceedings* 57–58 (2011).

[9] Tetlock, P. C. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* **62**, 1139–1168 (2007).

[10] Tetlock, P. C., Saar-Tsechansky, M. & Macskassy, S. More than words: quantifying language to measure firms' fundamentals. *The Journal of Finance* **63**, 1437–1467 (2008).

[11] Loughran, T. & McDonald, B. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance* **66**, 67–97 (2011).

[12] Pang, B. & Lee, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* **2**, 1–135 (2008).

[13] Barberis, N. & Thaler, R. A survey of behavioral finance. *Handbook of the Economics of Finance* **1**, 1053–1128 (2003).

[14] Da, Z., Engelberand, J. & Gao, P. The sum of all fears: investor sentiment and asset prices. *SSRN eLibrary* (2010).

[15] Singer, E. The use of incentives to reduce nonresponse in household surveys. *Survey nonresponse* 163–177 (2002).

[16] Boucher, J. & Osgood, C. E. The pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior* **8**, 1–8 (1969).

[17] Da, Z., Engelberand, J. & Gao, P. In search of attention. *The Journal of Finance* **66**, 1461–1499 (2011).

[18] Gilbert, E. & Karahalios, K. Widespread worry and the stock market **2**, 229–247 (2010).

[19] Granger, C. W. J. Investing causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969).

[20] Shapiro, K. *The Limits of Attention: Temporal Constraints in Human Information Processing* (Oxford University Press New York, 2001).

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

Acknowledgments

We would like to express our gratitude to trade-futures and Investor Intelligence for making their sentiment data available to our research program. We thank the Socionomics Institute for their kind advice, and their assistance in obtaining research data that was crucial in conducting the research presented in this paper. We also thank Professor Scott Smart, Fangzhou Liu of Kelley Business School at Indiana University, and Professor Pengjie Gao of Department of Finance at University of Notre Dame for their useful comments on this manuscript.

Author contributions

H.M., S.C. and J.B. developed the study, discussed the results, and contributed to the text of the manuscript. H.M. conducted the data analysis and generated figures.

Additional information

Competing financial interests: The authors declare that they have no competing financial interests.

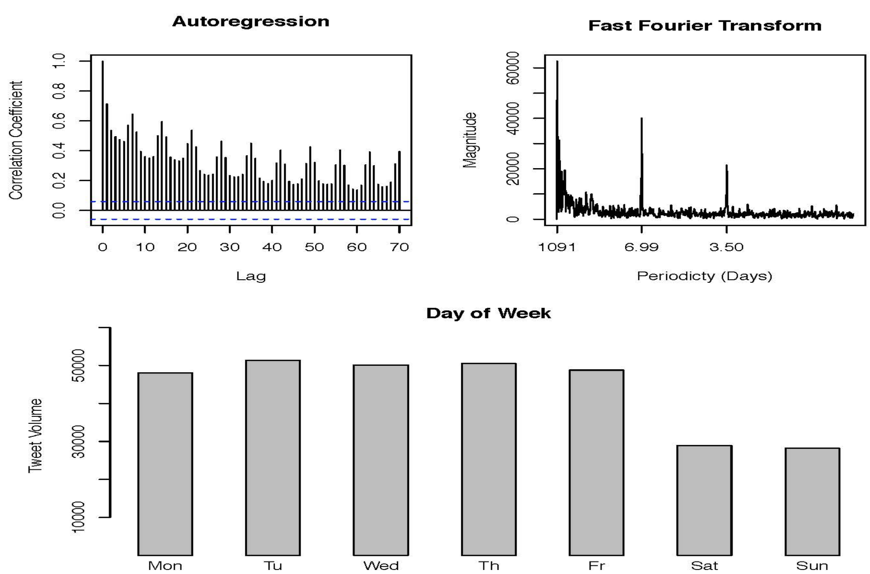
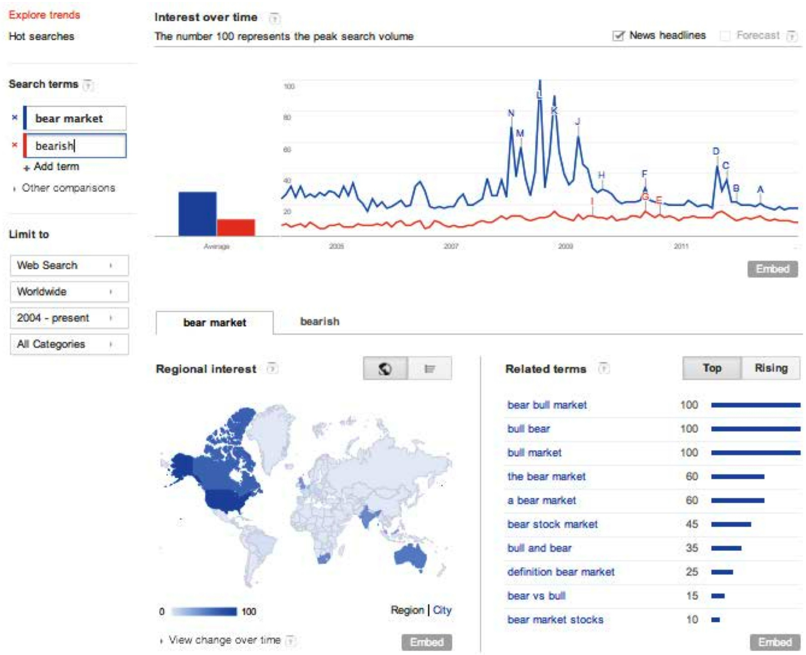


Figure 1: Bullish and Bearish Tweet Volume over Day of Week.



Google and the Google logo are registered trademarks of Google Inc., used with permission.

Figure 2: Google Trends with Search Queries “bear market” and “bearish”.

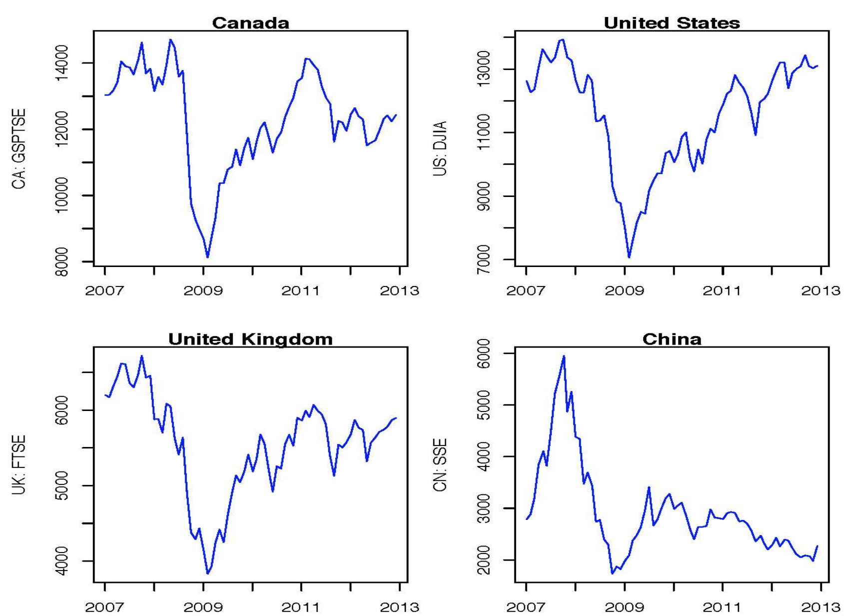


Figure 3: Monthly Stock Price of United States, United Kingdom, Canada and China.

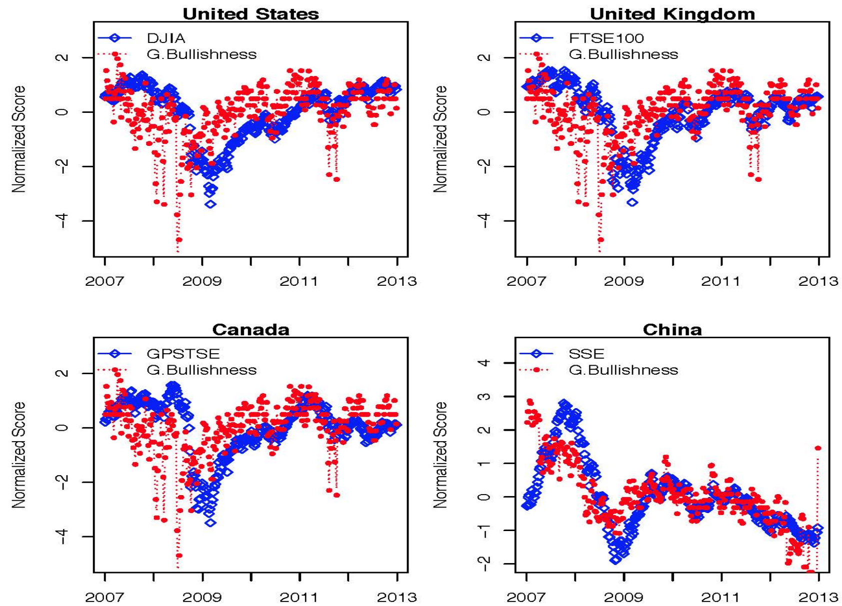


Figure 4: The Trend of Stock Market Price Against Google Bullishness.

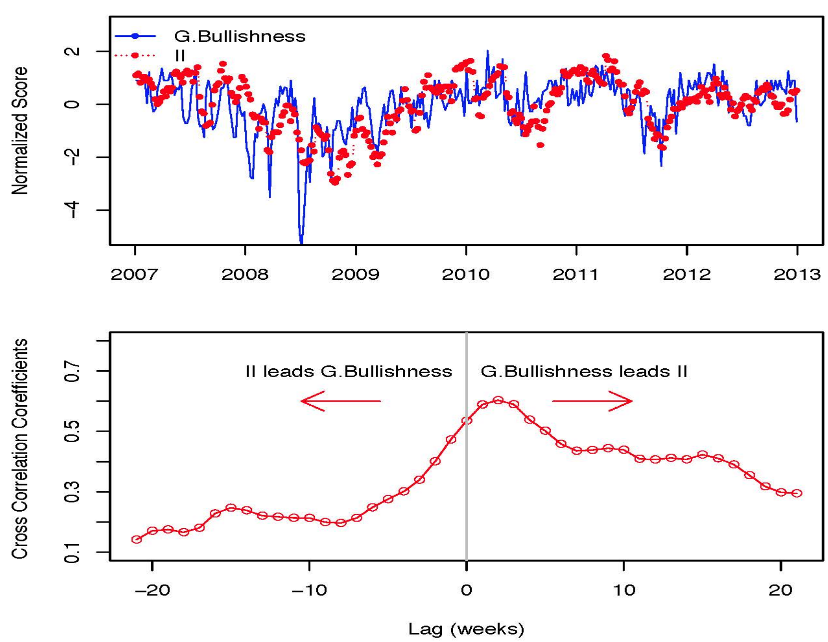


Figure 5: Correlations between Investor Intelligence (II) and Google Bullishness.

Table 1: Predicting Google Bullishness Using Twitter Bullishness

Bullishness	Coefficient	p
ΔG^B_{w-1}	**-0.54**	$\ll 0.01 \star\star\star$
ΔG^B_{w-2}	**-0.30**	$0.001 \star\star\star$
ΔG^B_{w-3}	-0.21	$0.02\star\star$
ΔG^B_{w-4}	0.009	0.91
T^B_{w-1}	**0.18**	$0.03 \star\star$
T^B_{w-2}	**0.09**	0.30
T^B_{w-3}	0.20	$0.03\star\star$
T^B_{w-4}	0.10	0.20
$p \leq 0.01$: $\star\star\star$, $p \leq 0.05$: $\star\star$, $p \leq 0.1$: $\star$		
Adjusted R^2=0.23, F=6.69 on df (8, 142), p$\ll$ 0.01		

Table 2: Predicting Daily Stock Returns of Dow Jones, S&P 500, Russell 1000 and Russell 2000 Using Twitter Bullishness.

Bullishness	DJIA		SP500		Russell1000		Russell2000	
Lag	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
1	**12.56**	$0.01\star\star\star$	**10.98**	$0.05\star\star$	**10.72**	$0.05\star\star$	**11.02**	$0.05\star\star$
2	2.27	0.67	2.61	0.65	2.46	0.67	2.66	0.65
3	2.18	0.69	3.69	0.53	4.037	0.48	4.58	0.43
4	-7.81	0.15	-8.10	0.16	**-9.99**	$0.08\star$	**-10.28**	$0.08\star$
5	-1.12	0.80	-1.28	0.79	-1.35	0.77	-1.37	0.78

Table 3: Predicting Stock Returns of US, UK, CA and CN Using Twitter Bullishness

Lag	US.DJIA		UK.FTSE		CA.GSPTSE		China.SSE	
	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
1	**13.18**	$0.01\star$	**17.98**	$0.0005\star\star$	**14.08**	$0.001\star\star$	**8.73**	$0.09\star$
2	1.30	0.81	**-10.39**	$0.06\star$	-5.26	0.26	-3.16	0.571
3	3.03	0.57	**11.11**	$0.04\star$	8.16	0.08	6.78	0.224
4	-8.79	0.10	**-9.85**	$0.07\star$	**-11.35**	$0.01\star$	-2.91	0.601
5	-2.31	0.60	-3.54	0.46	-1.799	0.64	-1.60	0.757

Table 4: Predicting Weekly Investor Intelligence Using Google Bullishness

Lag	II		G.Bullishness	
	Coeff.	p-value	Coeff.	p-value
1	0.08	0.18	**0.18**	$0.002 \star$
2	0.005	0.93	**0.19**	$0.002 \star\star$
3	-0.02	0.67	**0.19**	$0.003\star\star$
4	-0.06	0.27	0.002	0.98
Adjusted R^2 = 0.06, F=3.62 (df: 8 and 299), p=0.0005				

Table 5: Predicting Weekly Stock Returns Using Google Bullishness

Bullishness	US.DJIA	UK.FTSE100	CA.GSPTSE	CN.SSE
ΔG^B_{w-1}	-21.48 (0.24)	18.36 (0.36)	3.84(0.84)	4.91 (0.87)
ΔG^B_{w-2}	6.65 (0.73)	23.68 (0.27)	16.09 (0.44)	20.0 (0.53)
ΔG^B_{w-3}	-19.92 (0.29)	0.14 (0.99)	1.83 (0.93)	-16.39(0.60)
ΔG^B_{w-4}	-17.71 (0.34)	8.40 (0.67)	-7.07 (0.71)	-25.84 (0.38)
G^B_{w-1}	-24.38 (0.32)	33.8(0.26)	13.93 (0.64)	25.11(0.71)
G^B_{w-2}	35.87 (0.21)	9.26(0.78)	24.54 (0.46)	47.40 (0.54)
G^B_{w-3}	-30.24 (0.29)	-32.76(0.32)	-14.29 (0.66)	-63.20 (0.41)
G^B_{w-4}	18.28 (0.44)	8.14(0.78)	-2.80 (0.92)	18.99(0.77)

Outside and inside the parentheses "()" are regression coefficients and p-values, respectively.