EUROPEAN CENTRAL BANK

EUROSYSTEM

Statistics Paper Series

Huina Mao,
Scott Counts
and Johan Bollen

Quantifying the effects of online bullishness on international financial markets

No 9 / July 2015

Contents

Abstract **2**

1 Introduction **3**

2 Results **6**

2.1 Twitter bullishness 6

2.2 Google bullishness 7

2.3 International stock market 8

2.4 Research questions and inference 9

2.5 Lead-lag relationship between Twitter and Google bullishness 10

2.6 Twitter bullishness and stock market returns 11

2.7 Google bullishness and stock market returns 14

3 Conclusion **18**

Annex **19**

Methods 19

References **20**

Abstract

Computational methods to gauge investor sentiment from commonly used online data sources that rely on 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 based on Twitter updates and Google search queries. We examine the predictive power of this new investor bullishness indicator for 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 to investor sentiment and lead established investor sentiment surveys. The former, in particular, is a more powerful predictor of changes in sentiment in the stock market than the latter. Third, we observe that high Twitter bullishness predicts increases in stock returns, with these then returning to their fundamental values. We believe that our results may support the investor sentiment hypothesis in behavioural finance.

Keywords: computational science, investor sentiment, big data, social media, international financial markets

JEL codes: C1, C12

1 Introduction

According to the Efficient Market Hypothesis (EMH; Fama, 1970) investors operate as rational actors and share prices therefore fully reflect all existing, new, and even hidden information. Traditional efficient market models, however, fail to explain important market anomalies, such as the Great Crash of 1929, the Black Monday crash of October 1987, the dot-com bubble of the late 1990s and the stock market collapse of 2008. Behavioural finance challenges the EMH by emphasising the important role of behavioural and emotional factors in investor behaviour (Kahneman and Tversky, 1979; Shiller, 2006). Behavioural finance is based on two major assumptions: (i) "investor sentiment", i.e. the actions of investors are also determined by sentiment and not just rational considerations; and (ii) "limits-to-arbitrage", i.e. betting against irrational investors is costly and risky. Owing to the limited arbitrage of sophisticated investors, investor sentiment can influence stock prices (De Long et al., 1990a). In addition, investor sentiment or perceptions about the market can directly reflect general consumer sentiment about the economy, which can in turn influence consumer spending (Carroll et al., 1994). Knowing timely information on investor sentiment and consumer confidence can help government policy-makers and central banks to anticipate market trends and plan ahead. Therefore, the assessment and measurement of investor sentiment and its effects has become an important research topic (Baker and Wurgler, 2007).

In recent years, researchers have explored a variety of computational methods to measure the investor sentiment indicated by commonly used online data sources, such as stock message boards, news reports, microblogging environments, blogs and search engine queries. This approach holds considerable promise, given the unprecedented scale, high degree of detail, low cost and high frequency of the underlying data.

To the best of our knowledge, existing market sentiment measures are either classifier or dictionary-based. In Antweiler and Frank (2004), two popular classifiers – Naive Bayes and Support Vector Machine (SVM) classifiers – are employed to classify stock messages into three categories: "bullish", "bearish" and "neutral". This research has found that message bullishness and volume can help predict market volatility, but is of limited value when it comes to predicting returns. Similar results have been obtained in later work that uses as many as five classifier algorithms (Das and Chen, 2007). The latest and most relevant study (Oh and Sheng, 2011) classifies stock tweets from StockTwits® into the bullish and bearish categories, and builds a bullishness index that is shown to be predictive of future share price movements.

Together with machine learning approaches, a number of studies have focused on the development of linguistic lexicons or dictionaries to determine investor sentiment from the frequency of words in financial data sources. Perhaps the most influential study is that by Tetlock (2007), which uses the frequency of words on the Harvard negative word list (Havard-IV-4-TagNeg) in daily news items to construct a pessimism indicator; one found to predict the daily Dow Jones returns and company share prices reported in the author's later work (Tetlock et al., 2008). However, Loughran and McDonald (2011) argue that the Harvard Psychosociological Dictionary has been developed for the fields 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 comprised of 2,337 words, which was found to outperform the Harvard dictionary in measuring financial sentiment.

Classifiers and dictionary-based methods are useful for automatically processing large sets of text data used to produce general sentiment indicators. However, the variegated contexts and subtleties of human language pose a tough challenge for human raters and text analysis algorithms. In fact, the low accuracy with which humans themselves can assess text sentiment inevitably sets an unfavourable upper bound on what the best supervised classifiers can achieve. According to some studies (Das and Chen, 2007; Oh and Sheng, 2011; Pang and Lee, 2008), a machine learning classification accuracy of between 60% and 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. Furthermore, word-weighting schemes may be biased and context-sensitive, and dictionaries cannot be adjusted to reflect varying word contexts and semantics.

The limitations of automated sentiment analysis algorithms are not merely an academic or technical matter. Investors are averse to ambiguity and uncertainty (Barberis and Thaler, 2003). For computational indicators of investor sentiment to become an accepted part of the financial toolkit, they need to be reliable and accurate, and they also need to reduce ambiguity and risk rather than increase it.

In contrast to computational indicators, surveys of investor sentiment have already become an accepted part of the financial data environment. For example, the Daily Sentiment Index (DSI) and the weekly US Advisors' Sentiment Report issued by Investors Intelligence (II) involve two well-known surveys of investor sentiment. Since 1987, small traders have been interviewed about their bullish or bearish position on US futures markets in order to generate the DSI. And since 1963, II has investigated and categorised readers' opinions on market newsletters, assigning these to three categories, namely "bullish", "bearish" or "correction" (neutral). Simply put, surveys assess whether sentiment is bullish or bearish on the basis of the precise words used by people when responding to specific questions.

This certainly has the advantage of being unambiguous and exact, but surveys can still have some detrimental shortcomings, for example, they are resource intensive and expensive to conduct. Furthermore, what may seem a strength could actually be a weakness; when people are explicitly asked for their opinion, a variety of individual and social biases (including "groupthink") and the truthfulness of responses can become an issue (Da et al., 2010; Singer, 2002).

Here, we aim to define an indicator of investor sentiment that benefits from the advantage of traditional surveys by requiring explicit, unambiguous statements regarding investor sentiment, yet leverages large quantities of 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 occurrence of only two terms: “bullish” and “bearish” – chosen because they are rarely used other than in a financial context. Hence, they are more likely to produce an unambiguous indication of bullish or bearish investor sentiment.

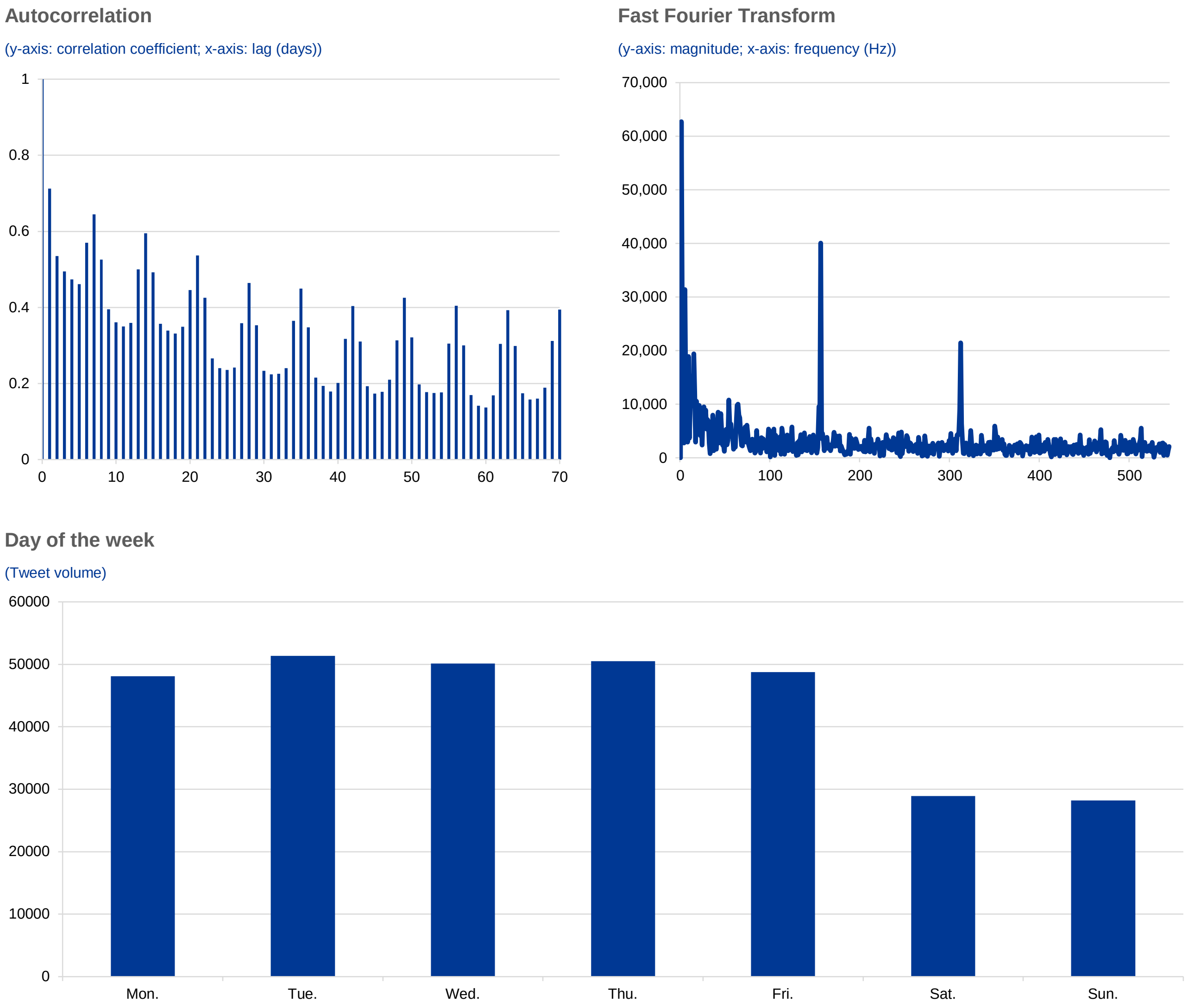
In this paper, we report the frequency of appearance of the terms “bullish” and “bearish” in Twitter content and Google queries over time, and define a bullishness index on the basis of their relative frequencies. We compare the bullishness indicators calculated respectively from Twitter content and Google queries, i.e. the same index is used for different data sources. We also compare both with existing surveys of investor sentiment, and examine their predictive effect vis-à-vis 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 and Google bullishness. Twitter bullishness has a statistically and economically significant predictive value in respect of share prices in the United States, the United Kingdom and Canada. We further observe that high Twitter bullishness indicates an increase in daily returns on the following day, with there being a return to normal levels within the next two to five days. Our results support the investor sentiment theory (De Long et al., 1990a), and suggest that Twitter bullishness may provide a useful and simple investor sentiment index.

2 Results

2.1 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 from 2010 to 2012, we find about 0.31 million bullish and bearish tweets. There are 1,091 days in total, and the average daily number of bullish and bearish tweets is 280. Figure 1 shows the volume of bullish and bearish tweets.

Figure 1
Daily volume of bullish and bearish tweets



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 (1,091 days) as the main periodicity, while the second and third ones appear at 6.99 days and 3.50 days respectively. Thus, 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 previous studies (Oh and Sheng, 2011) suggesting that the distribution of bullish or bearish messages matches investor behaviour. The average ratio of the number of bullish tweets to the total number of bullish and bearish tweets is 69.4%, indicating either a bias towards optimism on the part of online investors (Oh and Sheng, 2011) or the presence of a Pollyanna effect (according to the Pollyanna Hypothesis, humans universally favour positive words over negative ones; Boucher and Osgood, 1969).

In line with earlier work (Antweiler and Frank, 2004; Oh and Sheng, 2011), we define a Twitter bullishness index for which the value on day t is given by Equation (1).

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

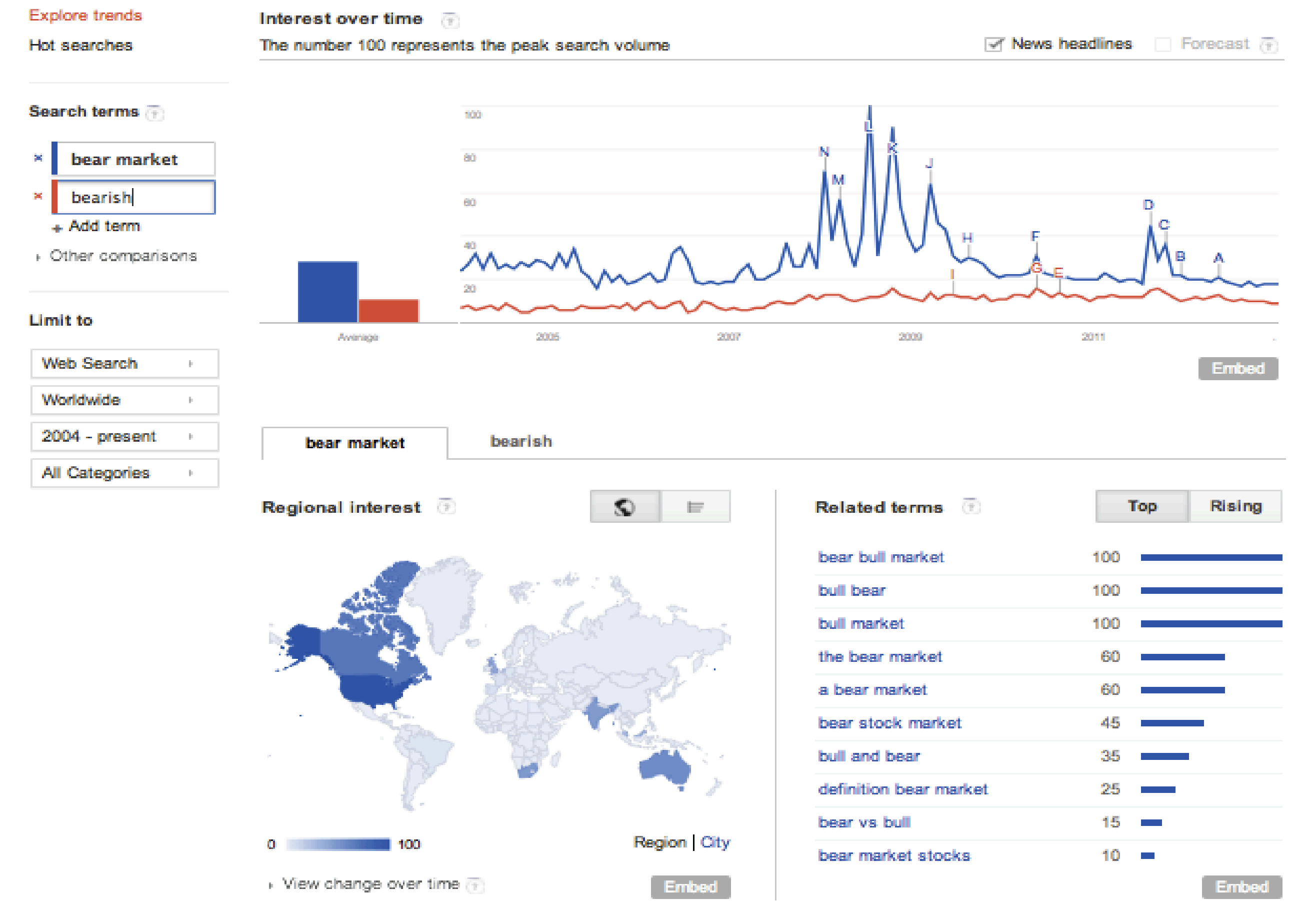
$\mathcal{B}_t$ and $\mathcal{R}_t$ denote the sets of bullish and bearish tweets on day t, respectively. The logarithmic transformation attenuates the effect of extremely large numbers of tweets. Studies have shown that this particular construction outperforms two alternatives (Antweiler and Frank, 2004).

2.2 Google bullishness

In a similar fashion to Twitter bullishness $\mathrm{T}_\mathrm{t}^\mathrm{B}$, namely as per Equation (1), we define Google bullishness $\mathrm{G}_\mathrm{w}^\mathrm{B}$ from the volumes of Google queries that contain the corresponding financial terms. The volume of such queries is determined using Google Trends, which necessitates a few notable changes. First, we find that the volumes of Google searches on the adjectives “bullish” and “bearish” are insignificant, most likely because isolated adjectives are rarely the subject of searches made by Google users. Indeed, Google Hot Trends indicates that the overwhelming majority of search queries are nouns. We therefore replace the adjectives “bullish” and “bearish” with equivalent terms, i.e. “bull market” and “bear market”, for our Google bullishness indicator $\mathrm{G}_\mathrm{w}^\mathrm{B}$. These ensure a greater depth of coverage (see Figure 2).

For China, we record the Mandarin ideograms 牛市 (bull market) and 熊市 (bear market). Second, Google search volumes are only available on a weekly basis whereas Twitter volumes are available for all points in time. Google bullishness $\mathrm{G}_\mathrm{w}^\mathrm{B}$ in week w is therefore defined in Equation (1) as the weekly ratio of $\|\mathcal{B}_\mathrm{w}\|$ and $\|\mathcal{R}_\mathrm{w}\|$, which represent the search volumes of “bull market” and “bear market” in week w, respectively.

Figure 2
Google Trends – search queries for “bear market” and “bearish”



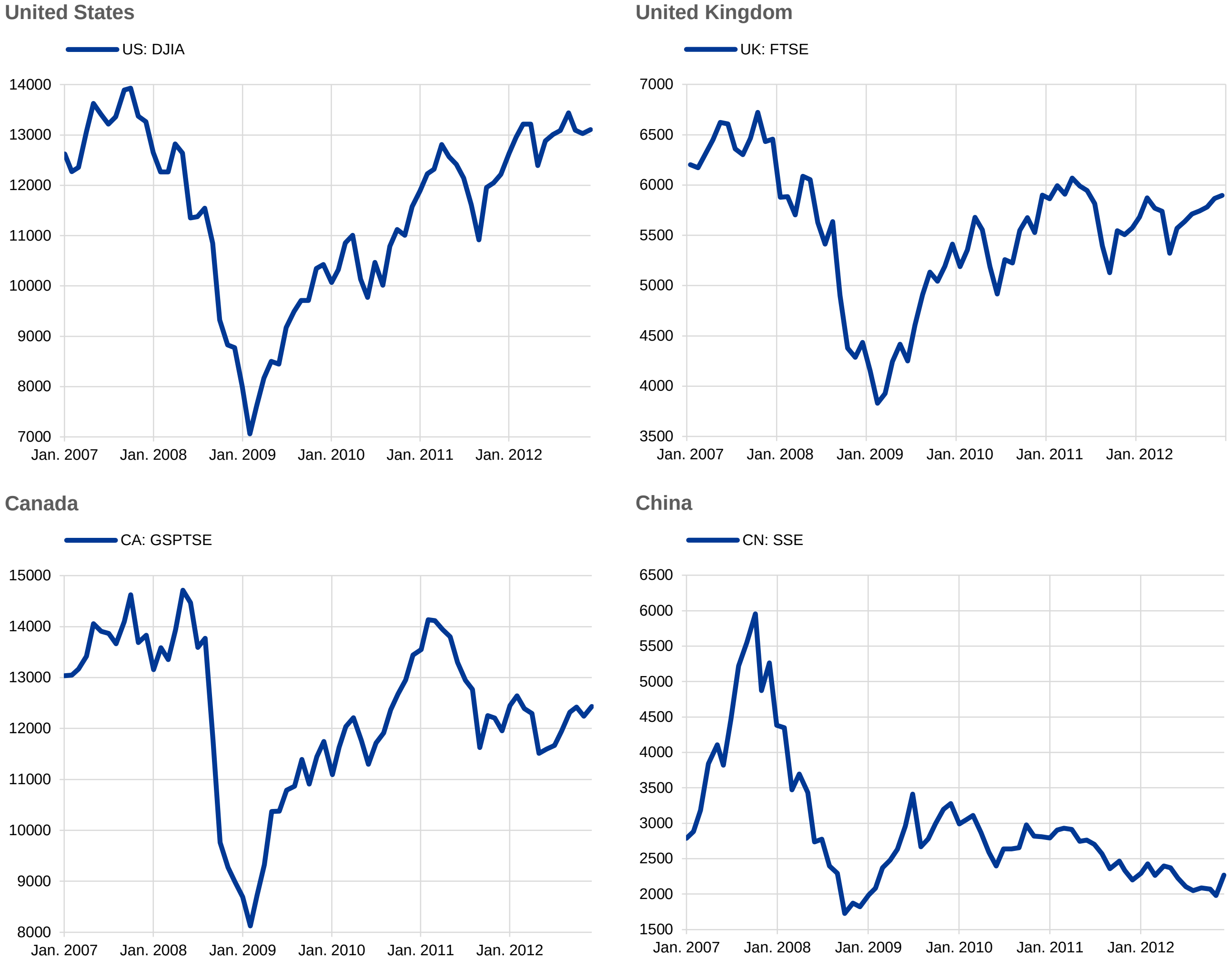
2.3 International stock market

In this paper, we compare Twitter and Google bullishness in relation to stock market values across four different countries (the United States, the United Kingdom, Canada and China) to increase the robustness of our results. These countries were selected for a number of reasons. First, their stock markets feature the largest market capitalisations in the world (i.e. according to World Bank statistics reported in 2012; see http://data.worldbank.org/indicator/CM.MKT.LCAP.CD). Second, both Google and Twitter are widely used in the United States, the United Kingdom and Canada. Therefore, online behaviour in these countries, as measured using Twitter and Google, is more likely to be representative of trends in the general population. Third, we deliberately include China in our study because its investor behaviour, market structure, legal system, as well as the uptake of social media and search engines, differ markedly from the other three countries. Hence, the country’s inclusion can help increase the diversity and robustness of our study.

Each nation's stock market is represented by a selected index, i.e. the Dow Jones Industrial Average (DJIA) for the United States, the FTSE 100 for the United Kingdom, the S&P/TSX Composite Index (GSPTSE) for Canada, and the SSE Composite Index (SSE) for China. Monthly prices of these four indices are shown in Figure 3.

Figure 3

Stock market indices in the United States, United Kingdom, Canada and China – monthly prices



2.4 Research questions and inference

We specifically set out to address three research questions. First, are Twitter and Google bullishness related? While the former is derived from daily microblogging updates on Twitter and the latter from weekly Google search query volumes, both originate from online activity and may thus 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 searches, does Google bullishness lead or lag weekly stock market returns? Throughout our prediction analysis, we control survey-based measurements of investor sentiment.

2.5 Lead-lag relationship between Twitter and Google bullishness

We compare Twitter bullishness and Google bullishness over time, and determine whether they are correlated, in particular whether one leads or lags the other.

Google bullishness $\mathrm{G}_{\mathrm{w}}^{\mathrm{B}}$ is a weekly time series while Twitter bullishness $\mathrm{T}_{\mathrm{t}}^{\mathrm{B}}$ is a daily time series; data on searches is only available from Google Trends on a weekly basis, whereas Twitter data can be collected for any time interval. In order to compare $\mathrm{G}_{\mathrm{w}}^{\mathrm{B}}$ and $\mathrm{T}_{\mathrm{t}}^{\mathrm{B}}$ for the same time period, we calculate the weekly mean of Twitter bullishness, denoted by $\mathrm{T}_{\mathrm{w}}^{\mathrm{B}}$. The sample period is thus the 156 weeks from 9 January 2010 to 29 December 2012.

We find a positive and statistically significant correlation between Twitter bullishness and Google bullishness (γ = +0.27, p = 0.0007). To estimate the lead-lag relationship between the two bullishness indices in both directions we use a vector autoregression (VAR) framework. This essentially involves a linear statistical model that captures the interdependencies among multivariate time series and is widely used to validate and quantify the predictability of financial indicators (Tetlock, 2007; Da and Gao, 2011; Gilbert and Karahalios, 2010). Our VAR model is equivalent to the bivariate Granger causality test (Granger, 1969) and is shown in Equation (2).

$$\Delta G_w^B = \alpha + \sum_{i=1}^{4} \beta_i \Delta G_{w-i}^B + \sum_{i=1}^{4} \chi_i T_{w-i}^B + \epsilon_w$$

The historical lag chosen is four weeks. Since VAR is sensitive to non-stationarity, we conduct an augmented Dickey-Fuller test, which indicates that $\mathrm{G}_{\mathrm{w}}^{\mathrm{B}}$ is non-stationary while $\mathrm{T}_{\mathrm{w}}^{\mathrm{B}}$ is stationary at a 90% confidence level. Therefore, we take the first order difference of Google bullishness, which is denoted by $\Delta\mathrm{G}_{\mathrm{w}}^{\mathrm{B}}$.

Table 1
Predicting Google bullishness using Twitter bullishness

Bullishness	Coefficient	p-value
$\Delta\mathrm{G}_{\mathrm{w-1}}^{\mathrm{B}}$	-0.54	<< 0.01***
$\Delta\mathrm{G}_{\mathrm{w-2}}^{\mathrm{B}}$	-0.30	0.001***
$\Delta\mathrm{G}_{\mathrm{w-3}}^{\mathrm{B}}$	-0.21	0.02**
$\Delta\mathrm{G}_{\mathrm{w-4}}^{\mathrm{B}}$	0.009	0.91
$\mathrm{T}_{\mathrm{w-1}}^{\mathrm{B}}$	0.18	0.03**
$\mathrm{T}_{\mathrm{w-2}}^{\mathrm{B}}$	0.09	0.30
$\mathrm{T}_{\mathrm{w-3}}^{\mathrm{B}}$	0.20	0.03**
$\mathrm{T}_{\mathrm{w-4}}^{\mathrm{B}}$	0.10	0.20

p≤0.001: ***, p≤0.05: **, p≤0.1: *
Adjusted R^2=0.23,F=6.69 on df (8, 142), p≤0.01

All variables in our regression model are normalised to standardised scores. Table 1 lists coefficient estimates with p-values. The reported coefficients measure the impact of a one standard deviation increase in an independent variable on the change in Google bullishness during week w. We find that ϵ_{w} satisfies the linear regression assumptions of independence, homoscedasticity and normality.

From Table 1, it can be observed that Twitter bullishness has a statistically significant and positive influence on the change in Google bullishness in the following week. But $\Delta\mathrm{G}_{\mathrm{w-1}}^{\mathrm{B}}$ and $\Delta\mathrm{G}_{\mathrm{w-2}}^{\mathrm{B}}$ are negatively related to the change in Google bullishness ΔG_w^B. We surmise that the negative sign may be the result of limited human attention spans (Shapiro, 2001), i.e. the focus of Google searches may move from one topic to another in the space of two or three weeks, depending on the attention span of users.

We note that only 23% of the variance of ΔG_w^B can be explained, indicating the difficulty of making predictions on the basis of sample data sources like Twitter and Google. In addition, when we reverse the regression direction, we do not find any significant prediction relationship running from ΔG_w^B to T_w^B. In other words, Twitter bullishness leads Google bullishness, but not vice versa. This finding may indicate a potential efficiency gain that gives Twitter an advantage over Google Search, but we leave it to future research to examine this issue further. Perhaps it can also explain the latter effect in more detail.

2.6 Twitter bullishness and stock market returns

Given that Twitter bullishness leads Google bullishness, we first apply the VAR model to examine whether Twitter bullishness has a predictive value in respect of stock market returns.

Table 2

Predicting the daily returns of selected stock market indices 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*****	**10.98**	**0.05****	**10.72**	**0.05****	**11.02**	**0.05****
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***	**-10.28**	**0.08***
5	-1.12	0.80	-1.28	0.79	-1.35	0.77	-1.37	0.78

We study the US stock market, which is the largest in the world. Furthermore, the United States has the highest concentration of Twitter users in the world. There are several major US market indices, including the DJIA, the S&P 500 (SP500) and the Russell 3000 Index. The aforementioned indices include the 30, 500 and 3,000 largest companies, respectively. Meanwhile, the Russell 3000 Index can be further divided into the Russell 1000, covering the top 1,000 companies (large-cap stocks), and the Russell 2000, covering the bottom 2,000 companies (small-cap stocks). To test the robustness of our method, we examine the Twitter bullishness prediction for all the major US stock indices.

The log stock return (R_t) is calculated on the basis of Equation (3)

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

where S_t^{close} and S_t^{open} are the stock market closing and opening prices on day t, respectively. Since daily Twitter bullishness is calculated from 00:00 to 23:59:59 Greenwich Mean Time (GMT) and daily US market returns are computed from 16:00

to 15:59:59 Eastern Time (ET; 21:00 to 20:59:59 GMT), the log return R_t is calculated from opening prices to closing prices on the same day t, rather than from closing prices on date t-1 to closing prices on day t. This can help avoid the possibility of including after-hours 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 research (Tetlock, 2007), the endogenous variables of our model include the share price as well as the trading volume in order to take into account liquidity effects. Log trading volume is de-trended to ensure stationarity. The third endogenous variable is our Twitter bullishness index Bt. The exogenous variables include VIX (a volatility index often referred to as the "fear index") and the Daily Sentiment Index (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:

$$R_t = \alpha + \sum_{i=1}^{5} \beta_i R_{t-i} + \sum_{i=1}^{5} \chi_i T_{t-1}^{B} + \sum_{i=1}^{5} \delta_i \mathrm{Vol}_{t-i} + \phi_i Exog_t + \epsilon_t$$

Table 2 shows the regression coefficient estimates and associated p-values. Each coefficient indicates the impact of a 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 a near absence of autocorrelation. In addition, ϵ_t in the model is found to have a normal distribution.

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

To examine the robustness of the predictive value of Twitter bullishness, we perform further tests against the large-cap SP500, the large-cap Russell 1000 and the small-cap Russell 2000 indices. The results of this analysis are reported in the second, third and fourth columns of Table 2, respectively. It is found that Twitter bullishness on the previous day has statistically and economically significant effects on these three indices. Moreover, we observe a price reversal for the four-day lag for all of the indices mentioned in Table 2, albeit a statistically insignificant one for the DJIA and the SP500. In particular, for the Russell 1000 and the Russell 2000, the initial increases on the first day are almost completely offset by the reversal on the fourth day (lag four). Our finding is consistent with the investor sentiment model (De Long et al., 1990a), where the irrationality of noise traders may cause an asset price to deviate from its fundamental value temporarily and then fall back to the mean.

In addition to the US stock market, we test the predictive value of Twitter bullishness on the stock markets of the United Kingdom, Canada and China. Twitter is widely used in the former two countries, so one may expect that Twitter bullishness may also contain relevant information for the UK and Canadian stock markets. Given that Twitter is not used in China, the comparison between Twitter bullishness and the Chinese stock market can serve as a null model, i.e. one would expect that Twitter bullishness has much less forecasting power for the Chinese stock market than in other countries. We use the VAR model to validate our assumptions.

Owing to the limited availability of existing predictive indicators for the UK, Canadian and Chinese markets, we adopt a reduced regression model in Equation (5) to examine the forecasting power of Twitter bullishness in relation to the stock markets of these countries.

$$R_t = \alpha + \sum_{i=1}^{5} \beta_i R_{t-i} + \sum_{i=1}^{5} \chi_i T_{t-i}^{B} + \epsilon_t$$

Daily returns are computed based on a country's main stock market index, namely the DJIA for the United States, the FTSE 100 for the United Kingdom, the GSPTSE for Canada, and the SSE for China. The regression coefficient estimates are reported in Table 3. The coefficient measures the impact of a one standard deviation increase in Twitter bullishness on daily returns in terms of basis points.

Table 3
Predicting stock market returns in selected countries 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*	17.98	0.0005**	14.08	0.001**	8.73	0.09*
2	1.30	0.81	-10.39	0.06*	-5.26	0.26	-3.16	0.571
3	3.03	0.57	11.11	0.04*	8.16	0.08	6.78	0.224
4	-8.79	0.10	-9.85	0.07*	-11.35	0.01*	-2.91	0.601
5	-2.31	0.60	-3.54	0.46	-1.799	0.64	-1.60	0.757

We find that both the reduced model of Equation (5) and the full model of Equation (4) generate nearly the same results regarding the predictive value of Twitter bullishness vis-à-vis the DJIA. A one standard deviation movement in Twitter bullishness causes a circa 13-basis points impact on the following day's Dow Jones 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 stock market prediction that is not captured by existing variables.

Furthermore, our results are robust. We find that Twitter bullishness has a similar predictive value for the UK and Canadian stock markets. We also observe similar “reversal” effects – these are stronger than in the case of the United States. As regards predicting developments in Chinese financial markets, we find that Twitter bullishness has a much lower predictive value (8.73 basis points) with only marginal statistical significance ($p = 0.09$). This is perhaps because Twitter has been shut down in China. Instead, Weibo is the most popular microblogging platform in this country.

2.7 Google bullishness and stock market returns

We obtain the search volumes for “bull market” and “bear market” from Google Trends for the period from January 2007 to December 2012, which constitutes 313 data points (weeks) in total. Google bullishness is calculated based on Equation (1). Figure 4 plots the trends in stock index prices and Google bullishness. We track the search volumes of “bull market” and “bear market” in both English and Chinese. Chinese Google bullishness is determined using the frequency of the ideograms 牛市 (bull market) and 熊市 (bear market) in user searches.

The Pearson linear correlation coefficients between Google bullishness and the corresponding log stock market index prices of the US, UK, Canadian and Chinese markets are 0.30, 0.38, 0.23 and 0.65 respectively; all being statistically significant ($p = 0.01$). Figure 4 shows the positive relationship between Google bullishness and index price levels. In addition, the former seems to lead the latter. Interestingly, this is particularly the case in extreme market conditions. For example, Google bullishness touched bottom in mid-2008, before the market turmoil of late 2008 and early 2009 in the United States, United Kingdom and Canada. Similarly, Chinese Google bullishness reached a peak in early 2007, preceding the market peak of early 2008. Subsequently, a declining trend of bullishness was followed by a downward trend in the market until 2009. It is surprising to find that Chinese Google bullishness has the highest correlation ($\gamma = 0.65$) with the Chinese market relative to the markets of the other three countries under consideration, where Google is the leading search engine: in China, Google accounted for less than 15% of the search market in 2012, compared with over 75% for Baidu, a local provider. The stronger positive correlation between the Chinese stock market and Google bullishness may be attributed to the country’s large number of Internet users (over 500 million in 2012). This result suggests that studying the value of online information sources for market prediction would be highly rewarding in the Chinese context; a topic that has received little interest in the literature.

Significant correlations between Google bullishness and stock market prices do not tell us whether one leads the other. Following the same regression framework adopted above, we investigate the extent to which weekly Google bullishness can predict market returns, i.e. the difference between the log closing price of this week and that of last week. However, both the level and change of Google bullishness are not predictive of the weekly returns of the Dow Jones, the FTSE 100, the GSPTSE and the SSE (see Table 4).

Table 4

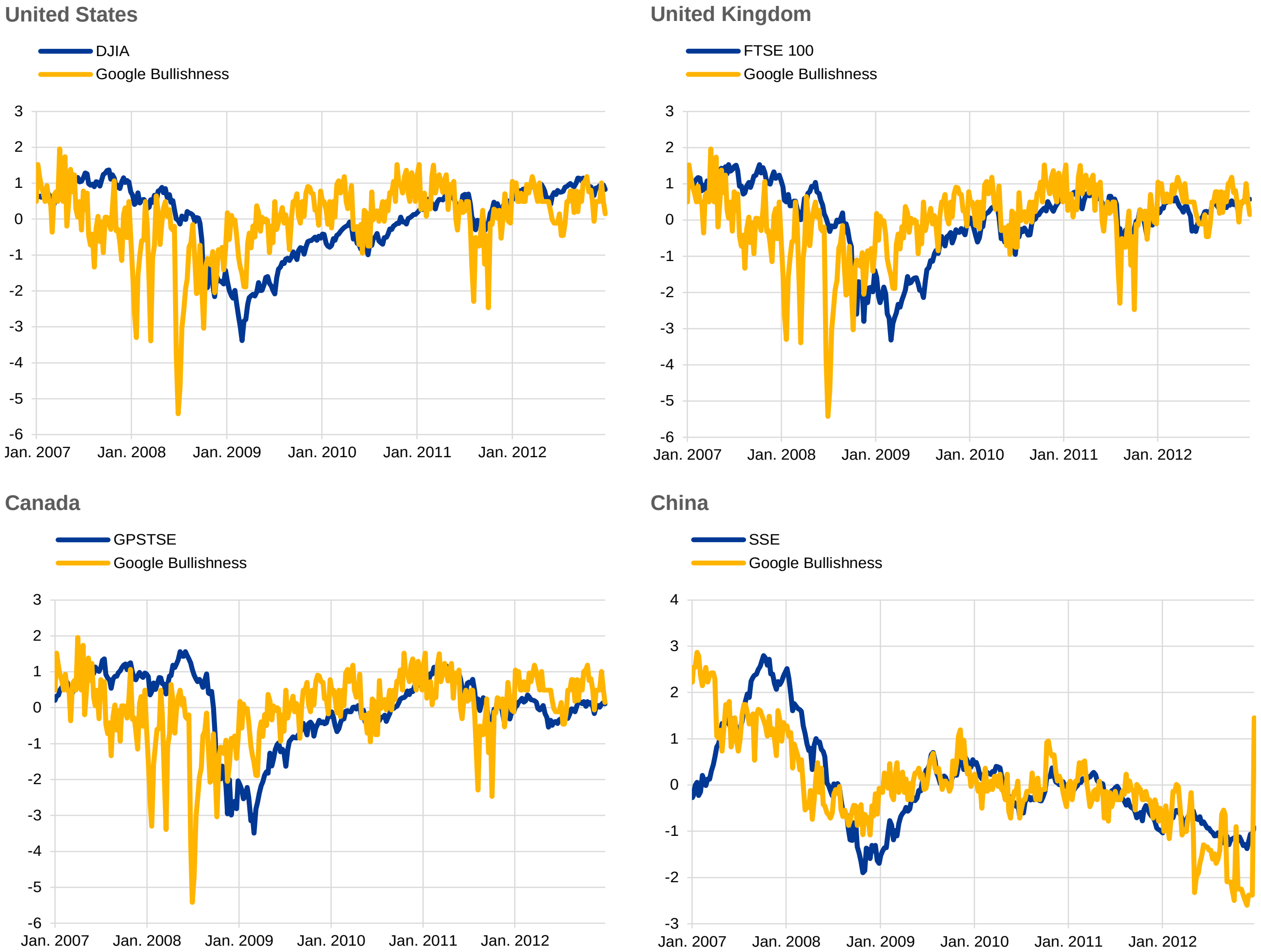
Predicting weekly stock index returns using Google bullishness

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)

Note: The numbers outside brackets refer to regression coefficients, while *p*-values are listed in brackets.

Figure 4

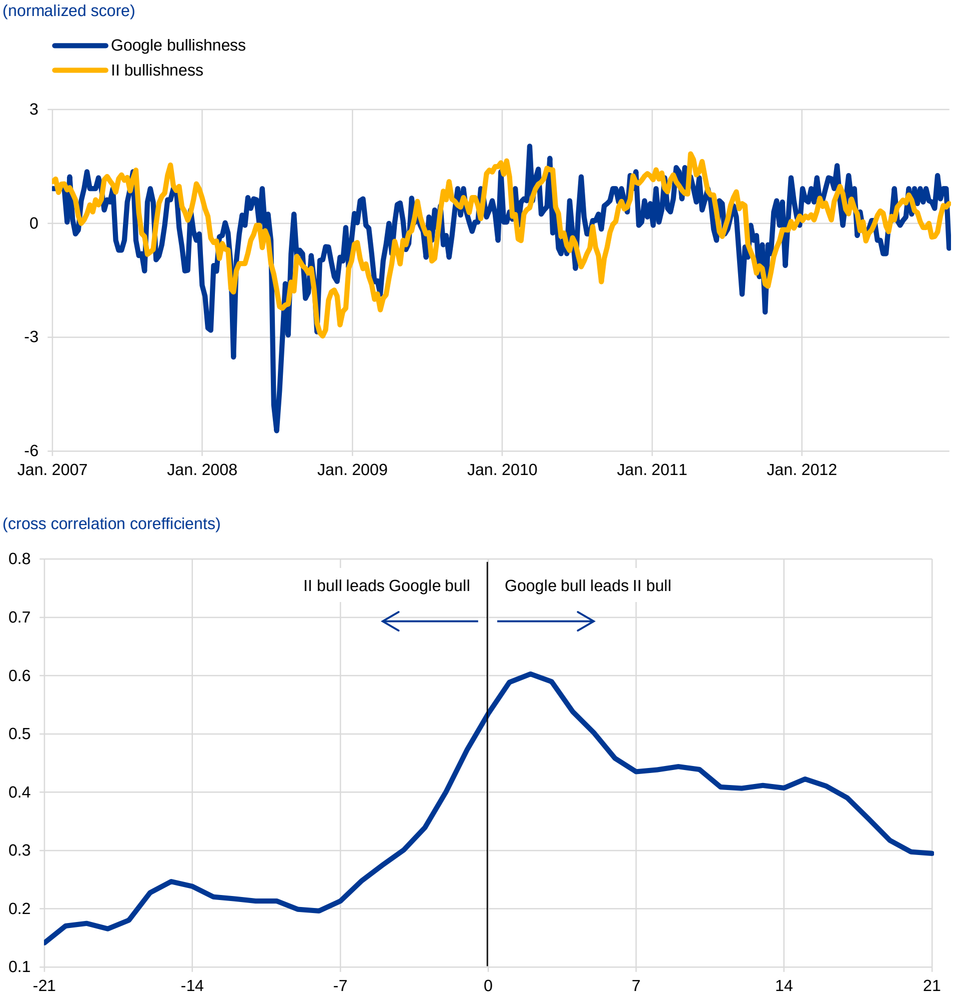
Trends in stock index prices and Google bullishness



We note that the lack of a predictive value of Google bullishness regarding the four stock markets under investigation may be explained by the fact that Google Trends data is provided only on a weekly basis. Over that time span, the market is likely to incorporate useful information and adjust prices accordingly, which means that Google bullishness derived from weekly Google Trends data would not contain predictive information.

Figure 5

Correlation between II bullishness and Google bullishness



In the opposite direction, we test the impact of weekly returns on the level and change of Google bullishness. The results are highly statistically significant. This finding supports the positive feedback trading theory in (De Long et al., 1990b), which states that traders' optimism increases when stock prices increase and their pessimism increases when prices decrease.

Despite the failure in predicting weekly index returns, we test whether Google bullishness may convey predictive information for investor sentiment rather than market prices. The US Advisors' Sentiment Report of Investors Intelligence is a well-accepted investor sentiment index in finance that measures whether US financial advisers are bullish, bearish, or neutral. Based on Equation (1), we compare "II bullishness" with Google bullishness: Figure 5 displays the general trends and their cross-correlation results. For lags in the range of minus three to three weeks, the correlation coefficients are 0.34, 0.40, 0.47, 0.54, 0.59, 0.60 and 0.59 (when moving in a positive direction).

The linear correlation between II bullishness and Google bullishness measured in the US case is highly positive: γ = 0.54, $p = 0.01$. More importantly, from the cross-correlation results in Figure 5, we observe that Google bullishness may in fact lead II bullishness. We use VAR to estimate the predictive relationship between these two different sentiment indicators. The time series are de-trended to be stationary by taking the first order difference. The results are shown in Table 5.

Table 5

Predicting weekly II sentiment using Google bullishness

Lag	II bullishness		Google bullishness	
	Coeff.	p-value	Coeff.	p-value
1	0.08	0.18	0.18	0.002
2	0.005	0.93	0.19	0.002
3	-0.02	0.67	0.19	0.003
4	-0.06	0.27	0.002	0.980

Adjusted $R^2 = 0.06, F = 3.62$, df (8, 299), $p = 0.0005$

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 one to three weeks. However, the regression model only explains about 6% of the variance, which suggests that it is difficult to predict changes in investor sentiment using these variables.

3 Conclusion

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 terms commonly used by investors in Twitter updates and Google search queries. Daily Twitter bullishness is indeed found to be a useful investor sentiment indicator. Our analysis shows a positive correlation between Twitter bullishness and Google bullishness on a weekly basis; furthermore, it finds 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 to existing surveys of investor sentiment, such as the Daily Sentiment Index and the US Advisors' Sentiment Report of Investors Intelligence. More importantly, we find that daily Twitter bullishness leads stock index returns in the United States (Dow Jones, SP500, Russell 1000, Russell 2000), the United Kingdom (FTSE 100) and Canada (GPSTSE), but has only very modest predictive value in respect of the Chinese stock market, as expected. While high Twitter bullishness predicts an increase in stock returns, we observe that these return to fundamental values within a week. Our research thus appears to support the hypothesised role of "investor sentiment" in behavioural finance. We also note the strong positive linear correlation between Google bullishness and Chinese stock index prices ($\gamma = 0.65$, $p = 0.01$), with the former apparently leading the latter in extreme market conditions. This result suggests the merits of studying the predictive value of online information sources such as Weibo for the Chinese market; a topic that has received little interest in the literature.

While our study shows a promising predictive correlation between Twitter bullishness and stock market prices, it offers no information with regard to causality. Causal inference is important for result interpretation, robust prediction and policy-making. Drawing causal inferences from Big Data is a challenging research problem. Consequently, future work will focus on developing a novel theoretical framework by combining experimental design methods and machine learning algorithms to infer the causal relationship between Twitter bullishness and financial markets.

Annex

Methods

Twitter and Google bullishness

We derive Twitter and Google bullishness scores from the volume of bullish and bearish tweets and related Google search queries. We simply select the words "bullish" or "bull market" and "bearish" or "bear market" to identify bullish and bearish sentiment as these are rarely used in a non-financial context and their meaning is relatively unambiguous. The definition of the online bullishness index is shown in Equation (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 period January 2010 to December 2012. Google search query data were retrieved from Google Trends (http://www.google.com/trends/) in 2012; this provides weekly search volume data on all queries made after January 2004. Values are dynamically scaled to the range of [0, 100], between volume peaks and troughs. Data from two investor sentiment surveys, the Daily Sentiment Index (http://www.trade-futures.com/dailyindex.php) and the US Advisors' Sentiment Report of Investors Intelligence (http://www.investorsintelligence.com/x/us_advisors_sentiment.html) were kindly made available to us. All historical market data were retrieved from Yahoo Finance! (http://finance.yahoo.com/) in 2012.

References

Antweiler, W. and Frank, M. Z. (2004), “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards”, *The Journal of Finance*, Vol. 59, No 3, pp. 1259-94.

Baker, M. and Wurgler, J. (2007), “Investor Sentiment in the Stock Market”, *Journal of Economic Perspectives*, Vol. 21, No 2, pp. 129-152.

Barberis, N. and Thaler, R. (2003), “A survey of behavioral finance”, in Constantinides, G.M., Harris, M. and Stulz, R.M. (eds.), *Handbook of the Economics of Finance*, 1st edn, Vol. 1, pp. 1053–1128, Elsevier.

Boucher, J. and Osgood, C. E. (1969), “The Pollyanna hypothesis”, *Journal of Verbal Learning and Verbal Behavior*, Vol. 8, pp. 1–8.

Carroll, C. D., Fuhrer, J. C. and Wilcox, D. W. (1994), “Does Consumer Sentiment Forecast Household Spending? If So, Why?”, *The American Economic Review*, Vol. 84, No 5, pp. 1397–1408.

Da, Z., Engelberg, J. and Gao, P. (2009), “The Sum of All FEARS: Investor Sentiment and Asset Prices”, SSRN eLibrary.

Da, Z., Engelberg, J. and Gao, P. (2011), “In Search of Attention”, *The Journal of Finance*, Vol. 66, pp.1461-99, October.

Das, S. R. and Chen, M. Y. (2007), “Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web”, *Management Science*, Vol. 53, No 9, pp. 1375-88

De Long, B., Shleifer, A., Summers, L.H. and Waldmann, R. J. (1990a), “Noise Trader Risk in Financial Markets”, *Journal of Political Economy*, Vol. 98, No 4, pp. 703-738.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990b), “Positive Feedback Investment Strategies and Destabilizing Rational Speculation”, *The Journal of Finance*, Vol. 45, pp. 379-95.

Desroches, B. and Gosselin, M.-A. (2002), “The Usefulness of Consumer Confidence Indexes in the United States”, *Bank of Canada Working Paper*, No 2002-22, August.

Fama, E. F. (1970), “Efficient capital markets: A review of theory and empirical work”, *The Journal of Finance*, Vol. 25, pp.383–417, May.

Gilbert, E. and Karahalios, K. (2010) “Widespread worry and the stock market” in *Proceedings of the fourth international AAAI conference on weblogs and social media*, Vol. 2, pp. 229-47.

Granger, C. W. J. (1969), “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods”, *Econometrica*, Vol. 37, pp. 424-38.

Kahneman, D. and Tversky, A. (1979), “Prospect Theory: An Analysis of Decision under Risk”, *Econometrica*, 47 (2), pp. 263–91, March.

Loughran, T. and McDonald, B. (2011), “When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-ks”, *The Journal of Finance*, Vol. 66, pp.35-65, February.

Oh, C. and Sheng, O. R. L. (2011), “Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement”, *ICIS 2011 Proceedings*, pp. 57-8.

Pang, B. and Lee, L. (2008), “Opinion Mining and Sentiment Analysis”, *Foundations and Trends in Information Retrieval*, Vol. 2, pp. 1–135.

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

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

Singer, E. (2001), “The Use of Incentives to Reduce Nonresponse in Household Surveys”, in Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (eds.), *Survey Nonresponse*, pp. 163-178.

Tetlock, P. C. (2007), “Giving Content to Investor Sentiment: The Role of Media in the Stock Market”, *The Journal of Finance*, Vol. 62, pp. 1139-68, June .

Tetlock, P. C., Saar-Tsechansky, M. and Macskassy, S. (2008), “More Than Words: Quantifying Language to Measure Firms’ Fundamentals”, *The Journal of Finance*, Vol. 63, 1437-67, June.

Acknowledgements
We would like to express our gratitude to trade-futures and Investors Intelligence for making their sentiment data available to our research programme. We thank the staff of the Socionomics Institute for their kind advice and their assistance in obtaining data that was crucial for conducting the research presented in this paper. We also thank Professor Scott B. Smart and Fangzhou Liu of the Kelley School of Business at Indiana University, as well as Professor Pengjie Gao of the Mendoza College of Business at the University of Notre Dame, for their useful comments.
An earlier version of this paper was presented at the European Central Bank workshop on "Using big data for forecasting and statistics", held in Frankfurt am Main in April 2014.

Huina Mao
School of Informatics and Computing, Indiana University, United States; e-mail: huinmao@indiana.edu
*The author currently works in Oak Ridge National Laboratory, US.

Scott Counts
Microsoft Research, Redmond, Washington, United States

Johan Bollen
School of Informatics and Computing, Indiana University, United States

Postal address	60640 Frankfurt am Main, Germany
Telephone	+49 69 1344 0
Internet	www.ecb.europa.eu

This paper can be downloaded without charge from www.ecb.europa.eu. Information on all of the papers published in the ECB Statistics Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/research/statistics-papers/html/index.en.html.

ISSN	2314-9248 (online)
ISBN	978-92-899-1905-0
DOI	10.2866/943096
EU catalogue number	QB-BF-15-001-EN-N