Social media bots and stock markets

Rui Fan (Swansea), Oleksandr Talavera (Swansea), and Vu Tran (Swansea)[†][✉]

This version: 30th April 2018

Abstract

This study examines how key stock indicators could be affected by information disseminated in social media networks by automated computer algorithms. We use a sample of about 49 million tweets with FTSE 100 firm names over 2015-2017. Our daily data findings suggest that such information is likely to affect volatility and trading volume of individual stocks. However, we do not observe significant influence of Twitter-bots on stock returns. Based on an intraday data, we detect the negative impact of tweets generated by bots on stock returns, but this effect vanishes within 30 minutes. The results for volatility and trading volume are consistent with the daily data analysis. The findings raise important policy implications for safeguarding financial stability and small investors' welfare.

Keywords: Social media bots, investor sentiments, noise traders, text classification, computational linguistics

JEL classification: G12, G14, L86

[†] We would like to thank Swansea research seminar and 2018 Royal Economic Society conference participants for helpful comments and suggestions. Any remaining errors are of our own.

[✉] Corresponding author: School of Management, Swansea University, Fabian Way, Swansea SA1 8EN, UK v.tran@swansea.ac.uk

Social media bots and stock markets

This version: 30th April 2018

Abstract

This study examines how key stock indicators could be affected by information disseminated in social media networks by automated computer algorithms. We use a sample of about 49 million tweets with FTSE 100 firm names over 2015-2017. Our daily data findings suggest that such information is likely to affect volatility and trading volume of individual stocks. However, we do not observe significant influence of Twitter-bots on stock returns. Based on an intraday data, we detect the negative impact of tweets generated by bots on stock returns, but this effect vanishes within 30 minutes. The results for volatility and trading volume are consistent with the daily data analysis. The findings raise important policy implications for safeguarding financial stability and small investors' welfare.

Keywords: Social media bots, investor sentiments, noise traders, text classification, computational linguistics

JEL classification: G12, G14, L86

1. Introduction

Enormous volume and velocity characterize information in social media such as Twitter where news transmits within milliseconds and the average daily volume is about five million messages. Recent rise of social media bots, cybots, social media farms (e.g. Ferrara et al., 2016) implies potential (temporary) dominance of intended misinformation in public domains. Demarzo et al (2003) show that repeated signals convert multi-dimensional to uni-dimensional (dis)agreements and are effective in polarizing public opinions. Meanwhile, rational inattentions may widely spread seeded misinformation among small agents due to costly (true) information acquisitions (see e.g. Sims, 2003). Automated social media bot accounts could be weaponized to influence political outcomes (see e.g. U.S Intelligence Committee, 2018) and to amplify fake news, manipulate stock markets (see e.g. Forbes, 2017). However, scientific evidence of such influences, to the best of our knowledge, is limited. We make a step further and explore the link between automated social media accounts and stock reactions. In particular, we investigate whether Twitter bots affect stock performance, volatility, trading volume and financial stability.

Our research is related to the literature exploring the interaction between stock markets and information in social media. Stock market participation and Twitter usage are positively correlated (Bonaparte and Kumar, 2013). Tweets can be used for forecasting aggregated market indexes. Bollen et al. (2011) find that some collective mood states derived from Twitter are linked to the value of the Dow Jones Industrial Average. Similarly, Zhang et al. (2011) show a significant relationship between the level of tweet emotionality and three major U.S. stock market indexes, namely Dow Jones, NASDAQ, and S&P 500. Among few works examining links between social media information flows and behavior of individual stocks, Sprenger et al. (2014a) find significant reactions from a S&P 500 company's stock prices to unusually high tweet volume on the company.[1]

To the best of our knowledge, prior financial literature ignores the fact that not all messages posted in social media are created by humans. Some Twitter users are bots, automated computer algorithms that are designed to pump intended information to public domains. Bots are used typically for sales promotions and brand awareness. However, bots could potentially be used to spread biased information which could ultimately influence human decisions (e.g. Gorodnichenko et al., 2017). For example, fake news stories on Hillary Clinton's poor health, weapons sales to Islamic extremists, or on Pope Francis's endorsement of Donald Trump are strongly associated with defections of the Democratic voters in Michigan (Howard et al., 2017). In 2013, fake news of an explosion that injured Barack Obama wiped out £130 billion stock value. Our paper fills this clear gap in prior literature and investigates how key stock indicators could be influenced by information spread by social media bots.

The contribution of our paper is fourfold. Firstly, our study is the pioneering attempt to investigate the impact of social media bots on stocks. While Gorodnichenko et al. (2017) argue that Twitter bots increase the polarization of public opinions during the 2016 US Presidential Election and the 2016 Brexit Referendum. We check whether the information flows generated by Twitter bots could have an impact on returns, volatility and trading volume of individual stocks. Secondly, prior works are based on daily stock prices while we employ an intraday stock data. Tweets can be posted and transmitted at an ultra-high frequency, it is thus necessary to zoom in and investigate under a high-frequency context. Microblog information posted in afternoons after market closes cannot affect the stock prices on that day. Thirdly, many studies argue that the sentiment of news is the most important element of its information content (e.g.

[1] Similarly, Sprenger et al. (2014b) show that there is a strong relationship between Twitter message sentiment, volume, and individual stock returns, trading volume and volatility. Ranco et al. (2015) show that cumulative abnormal returns react significantly to abnormal increases in tweeting activities, based on a sample of 30 companies from the Dow Jones Industrial Average.

Ryan and Taffler, 2004; Tetlock, 2007; Tetlock et al., 2008). We complement the existing literature and incorporate tweets' sentiments in our investigations of social media bots' impacts on multiple stock indicators including returns, volatility, and trading volume. This allows us to disentangle heterogeneous effects of social media bots on market participants. Final contribution relates to channel(s) of the impacts from social media information. Prior literature focuses on professional investing platforms while we aim at a much larger context. For instance, Sprenger et al. (2014a,b) examine tweets with ("$" and stock tickers). We argue potential impacts and contagion of (biased) beliefs could be via various channels including social communications (e.g. Hong et al., 2005). News without stock tickers or "$" in social media could affect firms' intangible assets e.g. United Airlines or Coca-Cola incidents in 2017.

Our main results suggest that information disseminated by Twitter bots increases volatility and trading volume of individual stocks regardless whether daily or high-frequency data is used. Although daily analysis does not detect any significant association between information generated by automated bot Twitter accounts and stock returns, we find that such information has a negative impact on stock returns at when stock prices are sampled at 5-minute frequency. The effect disappears within 30 minutes. In addition, we also use event study to detect abnormal increases in the volume of tweets and bot-tweets and find significant impacts on stock volatility, trading volume and liquidity. There is also a bounce-back pattern of price reactions in response to negative retweets, repeated information. It is noteworthy that human users could retweet information generated by automated bot Twitter accounts (see e.g. Ferrara et al., 2016).

Our study makes a multidisciplinary contribution to the field of social media bots, politics and stock market. Gorodnichenko et al. (2017) find significant spillover effects from bots to human activities on social media during Brexit referendum and U.S. election. A direct implication is that social media bots could influence political beliefs. However, it is difficult and very costly to measure political beliefs at a reasonable frequency for empirical investigations. We employ liquid stocks as a reasonable measure of social media bots' impacts due to a good number of bot-related events.

Our findings raise a number of important policy implications. Firstly, there are urgent needs for regulatory precautions to safeguard financial stability and small investors' welfare from the spread of misinformation in social media. Importantly, transparency of advertisements displayed on social media should be enhanced. Advertisements should be flagged as paid contents and indicate payers. Moreover, authorities should establish a code of practice and regulations to monitor social media providers such as Facebook and Twitter. Social media providers are more able and responsible to fight against the spread of fake information. Additionally, there could be problems arising from a lack of social media literacy. Nowadays, people are overwhelmed by huge amount of information in social media. It is costly (in terms of time and efforts) to determine the correct information. Lastly, our study suggests potential benefits of data availability for researchers. The large amount of data from social media network should be made available for investigations and studies which in turn play an important role in monitoring (ab)usages of social media networks. Investigations based enormous data in social media also could yield important insights for public welfare and relevant policy implications.

This paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the Twitter and stock data employed in the study. Section 4 explains the methodology. Section 5 presents the empirical results. Section 6 summarizes the findings and concludes.

2. Related literature

2.1 News, stock message boards and Twitter as a form of information intermediary

Business press, online stock message boards and Twitter all have been treated as an intermediary for information propagation. Sprenger et al. (2014b) argue that Twitter is a more comprehensive real time news database in two aspects. Firstly, Twitter users can post a tweet any time, while only a small number of authors publish news articles at discrete time in the business press. Secondly, traditional press usually only focuses on big news stories, but Twitter messages can deliver information about minor news stories. And the existing literature put little emphasis on minor news events. Additionally, Sprenger et al. (2014b) report that the stock microblogs have more information content than stock message boards in two ways. Firstly, message boards group postings into a separate board for every firm, hence some outdated information may still attract investors' attention, while Twitter presents more up to date information. Secondly, on stock message boards each posting is usually related to a specific stock, while Twitter users can be exposed to information for all stocks.

Several studies have examined the relation between news and stock returns. Dougal et al. (2012) find that there is a casual relation between the Wall Street Journal columnists and Dow Jones Industrial Average daily returns. Additionally, Engelberg and Parsons (2011) argue that local media coverage of earnings announcements of S&P 500 index firms can help to forecast local stock trading. Moreover, Kerl and Walter (2007) also show that German Personal Finance Magazines publish recommendations that help investors to earn positive abnormal returns within five days of the publication day. Furthermore, Chen et al. (2014) show that both the articles on the most popular investor social network platforms in the U.S. and the commentaries in response to the articles can predict future stock returns and earnings surprises. And Gross-Klussmann and Hautsch (2011) use a high-frequency VAR model and 20 seconds price data of UK stocks, and find clear responses in returns, volatility, trading volume and bid-ask spreads after news announcements.

Many researchers also pay attention to the large amount of qualitative user-generated information from the online stock forums, and the evidence are mixed. Wysocki (1998) investigates stocks listed on Yahoo! message boards and shows that message posting volume can forecast next day stock trading volume and return. Contrarily, Tumarkin and Whitelaw

(2001) examine internet service sector stocks and find that, consistent with market efficiency, online message board activity cannot predict industry adjusted returns or abnormal trading volume.

Another relevant study by Antweiler and Frank (2004) uses Naive Bayes algorithm to study the message information from both Yahoo!Finance and Raging Bull for 45 companies in the Dow Jones Industrial Average and the Dow Jones Internet Index. Antweiler and Frank (2004) show that stock message volume can help forecast volatility, while the effect on stock returns is statistically significant but economically small. Sprenger et al. (2014b) criticize that the sample period in 2000 coincides with the internet bubble and one third of their sample firms are the most affected technology companies. Another limitation of Antweiler and Frank (2004) is that real returns are used rather than abnormal returns, because Tumarkin and Whitelaw (2001) argue that it is important to separate buy and sell signals by using market-adjusted abnormal returns instead of real returns.

Our study explores the relation between key stock indicators and information dissemination in social media. We have a sample of about 49 million tweets with 55 FTSE 100 firm names during the period between July 2015 and August 2017. In particular, we examine whether automated information posted by Twitter bots could potentially affect returns, volatility and trading volume of individual stocks.

2.2 Examine the relations between tweet features and market features

Pritamani and Singal (2001) use stock price abnormal returns to separate between good and bad news. But their measure is then endogenous and may not provide the information content of the news. Ryan and Taffler (2004) also argue that if buy and sell signals are not separated, the impact that good and bad news have on stock prices may cancel out each other and hence there will be little effect on the market on aggregate.

Many studies show that there is an association between message bullishness and stock returns.

Hirshleifer and Teoh (2003) show that different forms of firm's information disclosures affect investors differently due to limited attention and processing power. Similarly, Barber and Odean (2008) argue that most investors only think about buying stocks that can catch their attention, because it is difficult for many investors to analyze every stock. Both Ng and Wu (2006) and Hong et al. (2005) also find that the trading decisions of investors are influenced via word of mouth.

Investors may also post more messages about the stocks that they trade. Wysocki (1998) and Antweiler and Frank (2004) both show that online stock board message volume can help to predict next day stock trading volume. The volume of messages posted on the online stock message board may also potentially affect the stock returns. Dewally (2003) finds that there are more buy recommendations than sell advice on the online stock message boards, with a ratio of greater than 7:1. While Antweiler and Frank (2004) argue that the effect of online stock message volume on stock returns is statistically significant but economically small.

Many studies find that online stock board message volume can be related to stock return volatility. Antweiler and Frank (2004) show that online stock board message volume can help forecast volatility. And Danthine and Moresi (1993) argue that more information lowers volatility as rational agents are better positioned to counteract the actions of noise traders. But Brown (1999) demonstrates empirically that noise traders may actually increase volatility. Koski et al. (2004) also show that online stock message board postings can increase volatility, although they also add that the reversed causation actually dominates. De Long et al. (1990) argue that the noise traders earn higher expected returns and face excess volatility than rational investors, as the unpredictability of noise traders' beliefs creates a risk that deters rational arbitrageurs from profit.

Our study finds a significant relationship between the automated information flows created in Twitter (bots), and volatility and trading volume of individual stocks. The findings based on high-frequency data for volatility and trading volume are compatible with the daily data results. However, we do not find any significant relations between Twitter-bots and stock returns. When we use intraday data, we observe the negative impact of tweets generated by bots on stock returns, but this effect disappears within 30 minutes.

2.3 Two theories about media coverage and investors' attention

Solomon et al. (2012) argue that there are two theories about how media coverage influences investors' decisions: the information view and the salience view. The information view states that media coverage can lower the cost of information acquisition and reduce information asymmetry between firms and investors (Tetlock, 2010). Specifically, Bushee et al. (2010) find that greater media coverage can reduce information asymmetry around earnings announcements. In a study more closely related to ours, Blankespoor et al. (2014) also show that the use of Twitter by firms to disseminate information can result in a reduction in information asymmetry.

While the salience view argues that more media coverage may bring the company more investors' attention and investments. (Solomon et al., 2012) For instance, Da et al. (2011) suggest Search Volume Index to better capture investor attention. They find that increasing attention can forecast high stock prices in the next two weeks and big first-day returns after IPOs. Solomon et al. (2012) examine mutual fund holdings and state that fund holdings with stocks recently covered in the press attract more investments than fund holdings with stocks not featured in the media.

Sprenger et al. (2014a) argue that it is not clear whether Twitter could reduce information asymmetry or capture investors' attention on specific stocks. A combination of both views may fit better. If the company is frequently covered by the media, the information view may apply. But if the firm is not much featured in the press, the salience view may better explain.

3. Tweet features and bot identifications

3.1. Tweet data collection and cleaning

This study uses Twitter Streaming application programming interface (API) to collect the Twitter data. API can be treated as interface between users and the system. The interface passes the enquiries raised by users to the system and then returns the answers to the users. We made requests to collect tweet messages with certain keywords and all tweets containing the keywords are collected in the harvest periods. Each tweet collected has information regarding the text of the tweet, user name and ID, date, location, friend and follower counts, etc. We collect 49.17 million tweets containing the name of a FTSE 100 company for the two-year period from 1st August 2015 to 31st July 2017.[2] Daily stock prices for FTSE 100 companies are collected from Datastream for the period from 1st January 2014 to 31st July 2017 to estimate abnormal measures of stock returns, and 5-minute stock prices are obtained from Bloomberg between February and July 2017. We examine FTSE 100 companies to get a representative sample of all UK equities, but focus on 55 well-known firms which have an average daily tweets of 100 or more. We could not obtain satisfactory Twitter data for some companies, e.g. Aberdeen Asset Management, as people do not like to tweet long company names due to Twitter's 140-character limit. We have tried alternative ways using either the ticker or dollar sign followed by the ticker to collect Twitter information, but we end up with too much noise for some firms, e.g. British Petroleum ('BP', '$BP'). A full list of 55 companies employed is reported in Appendix A1.

Following Gorodnichenko et al. (2017), we clean the tweets in three steps. First, we delete special characters in tweets such as link tokens (starting with 'http', 'https', 'www'), hashtag tokens (starting with '#'), and user identifier tokens (starting with '@') from the tweet

[2] We focus on the companies which have been included in FTSE 100 as of 1st January 2014.

messages. Second, all tweets containing only links or URLs are deleted. Finally, all non-English tweets are excluded.

3.2. Sentiment of Twitter messages

As sentiment of news is one of the most important features of its information content, our study separates positive tweets from negative tweets. We employ TextBlob, a text-processing tool in Python, to get a polarity score for each tweet posting. The polarity scores range from -1 to 1: a negative score indicates that the sentiment is negative, while a positive score means that the sentiment is positive, and 0 score suggests that the sentiment is neutral. Both PatternAnalyzer and NaiveBayesAnalyzer in TextBlob are employed to perform sentiment analysis, and the same sentiment score are obtained for each tweet posting, which supports the robustness of our sentiment analysis. Appendix B gives examples on how the polarity scores are obtained by TextBlob in our sample.

3.3. Humans vs. automated bots Twitter accounts

There is no perfect produce to recognise bot tweets as bots/cybots can imitate human behaviour. We identify Twitter bot accounts based on two techniques. Firstly, a Twitter account is classified as a bot account if we record suspicious activities in more than 50% of days during the sampled period. For example, if an account has tweeting activities on ten days, it is detected as a bot account if the following criteria are satisfied on more than five days. Three criteria for a suspicious activity are employed including (i) abnormal tweeting time, i.e. from 0:00 to 6:00 am; (ii) large number of tweets a day; (iii) repeating the same tweet content three times or more on one day. Secondly, we employ Botometer, an online social media bot detection tool developed by researchers from Indiana University and Northeastern University. Investigation results based on both techniques are consistent with each other.

3.4. Original tweets and retweets

We also separate between original tweets, the tweets that are posted for the first time, and their retweets. Firstly, the text of each tweet is checked and a new variable RT is generated. RT is 1 if the tweet begins with 'RT @', which means that this is a retweet; or 0 otherwise, which indicates that this is an original tweet. We then examine the content after '@' but before the main text and denote it as RT_from, which is the user name of the Twitter account from which the tweet was retweeted. Consequently, we can identify the original tweets, their direct and indirect retweets.

4. Aggregation of tweet information and estimation framework

4.1. Tweet postings aggregation

All daily tweets are aggregated to examine the relation between all tweet postings and stock price changes on a daily basis. We explore the relationship between market features (stock returns, trading volume, and volatility) and tweet features (positiveness, message volume, and agreement). We follow Antweiler and Frank (2004) and Sprenger et al. (2014b) to define tweet sentiments as

$$Positiveness_t = \ln\left(\frac{1 + M_t^{positive}}{1 + M_t^{negative}}\right) \quad (1)$$

where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day *t*. And tweet message volume is given as

$$Message_t = \ln(M_t)$$

where M_t is the count of all tweet messages containing one sample company name on day *t*. Tweet agreement is defined as

$$Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{positive} - M_t^{negative}}{M_t^{positive} + M_t^{negative}}\right)^2} \quad (2)$$

If all tweet messages are positive or negative, the agreement among all messages equals 1. The tweet volume, positiveness and agreement are defined as 0 if there is no tweet for a particular company on day *t*. We have tweet and stock features data available for most company-day observations. The problems caused by quiet periods i.e. non-tweet observations can be ignored. Finally, consistent with the UK trading hours (8:00 am to 4:30 pm), Twitter messages posted on and after the market closes at 4:30 pm are assigned to the following day.

4.2. Stock indicators

Following prior literature (e.g. Antweiler and Frank, 2004), we calculate log return from stock prices and define abnormal return as

$$AR_{i,t} = R_{i,t} - E(R_{i,t})$$

$R_{i,t}$ is the log-return for stock i on day t. $E(R_{i,t})$ is the expected return. We employ two alternatives for the expected return. First, the mean return during the previous 100 trading days between day -110 and day -10[3], thus

$$AR_{i,t} = R_{i,t} - \frac{1}{100}\sum_{k=10}^{110} R_{i,t-k}$$

The stock's idiosyncratic risk, however, cannot be reflected by the above mean-adjusted return, thus we estimate the expected return using the market model estimated by ordinary least squares (OLS) regressions

$$E(R_{i,t}) = \alpha_i + \beta_i(R_{m,t}) + u_i \quad \forall t \in \{1, 2, \dots T\}$$

and α_i is the intercept, β_i is the relation between the stock returns and market returns, u_i is the error term and T is the number of periods in our sample. Following previous literature, we adopt a 100-day estimation period starting 10 days prior to the relevant date.[4]

---

[3] We use an alternative 1-year estimation period, i.e. [-10,-260], and obtain qualitatively similar results. (available upon request)

[4] Similarly, an alternative 1-year estimation period yields qualitatively similar results. (available upon request)

The cumulative abnormal return is given as

$$CAR_{i,t} = \sum (AR_{i,t})$$

and the average cumulative abnormal return for N firms is

$$ACAR_t = \frac{\sum_{i=1}^{N} CAR_{i,t}}{N}$$

Parkinson (1980) estimated daily volatility using intraday high and low stock prices $S_{t,high}$ and $S_{t,low}$, we employ this measure which is defined as

$$Vol^{Park} = \frac{\ln\left(S_{t,high}/S_{t,low}\right)}{2\sqrt{\ln 2}}$$

We use both the Parkingson volatility measure and abnormal changes in the volatility measure in order to obtain robust findings. An abnormal change in the volatility measure equals today volatility minus the average volatility in the past 100 trading days, i.e. [-110, -10]. Lastly, trading volume is defined as the natural logarithm of the number of shares traded on a given day. We also employ a similar measure of abnormal changes in trading volume as above.

4.3. Summary statistics

There are in total about 49.17 million tweets containing the name of a FTSE 100 company; the average number of daily tweets is about 2,000, and the standard deviation is around 4,900 tweet postings per day. The large number of Twitter messages per company per day indicates that our sample comprises a sound information flow. Figure 1 plots the daily numbers of different types of tweets on the 55 sampled firms from FTSE 100 composites. The numbers vary vastly across time.

Table 1 presents the descriptive statistics of market and tweet features. Panel A of Table 1 describes the whole sample, and Panel B compares characteristics between high-bot activities and low-bot activities subsamples. We define high-bot and low-bot subsamples based on the proportion of bot-tweets over the total number of tweets on a day. If the proportion is larger than 50%, the observations on that day are included in the high-bot subsample, and vice versa. The size of low-bot subsample is much larger than that of high-bot subsample. Notably, firms in the high-bot subsample have significantly different key financial characteristics compared to companies in low-bot subsample. For example, earnings per share, profitability (i.e. ROE), leverage (debt over equity ratio) of high-bot subsample are significantly higher than those of low-bot subsample, while price per earning, ex-post daily return are significantly lower. It is noteworthy that the average daily return for the whole sample is 0.01%, and -0.14% for the high-bot subsample, which are equivalent to 2.5% and -35% annual returns. Additionally, positiveness in the high-bot subsample is significantly lower than that of the low-bot subsample, suggesting that there are more negative bot-tweets during high-bot activities periods.[5]

4.4. Empirical specification

We then present the regression specifications as

$$y_{i,t} = \alpha_i + \beta_1 Positiveness_{i,t} + \beta_2 Message_{i,t} + \beta_3 Agreement_{i,t} + \delta R_t^{FTSE} + \varepsilon_{i,t} \quad (3)$$

where i stands for firm and t denotes time, and $y_{i,t}$ is a vector of dependent variables, i.e. individual stock returns, volatility and trading volume.

We define $Positiveness_{i,t}$ variable using equation (1) to examine the impact of sentiment on stock market features. To note that we use tweets between 4:30pm, when the London Stock exchange closes on the previous trading day, and 4:29pm today to eliminate potential endogeineity. Consistent with Sprenger et al. (2014b), we anticipate a positive (negative) coefficient of Positiveness in explaining returns (volatility, trading volume).

$Message_{i,t}$ is defined as the natural logarithm of the number of tweets, between 4:30pm on the previous trading day and 4:29pm today. In line with Antweiler and Frank (2004) and Sprenger et al. (2014b), we anticipate a positive coefficient of Message to explain volatility

[5] The correlation matrix (reported in Appendix A1) shows significant correlations between tweet features and trading volume and volatility.

and trading volume, but not returns.

$Agreement_{i,t}$ describes the extent to which tweets agree to or are different from each other, i.e. similar or very different numbers of positive versus negative tweets. We use equation (2) to estimate Agreement. Compatible with Sprenger et al. (2014b), we anticipate negative coefficients of Agreement in explaining returns, volatility and trading volume.

Lastly, as the volume of tweets have substantial cross-sectional variations, we control for market return (FTSE 100 return) R_i^{FTSE} and firm individual fixed effects.

5. Results

5.1. Relation of tweet and market indicators

Table 2 reports fixed effects panel regressions of three different return measures as dependent variable and four tweet features (positiveness, volume, agreement and bots) for tweets containing a FTSE 100 firm name as independent variables. Mean-adjusted return is log return on a given day minus the average return over a 100-day estimation period beginning 10 days prior to the relevant date. Market-model return is log return on a given day minus the expected return estimated by the market model using FTSE 100 index as the market portfolio using the same estimation period as mean-adjusted return. Market return (FTSE 100 index return) is employed as a control variable. Given that the number of Twitter messages varies considerably across companies, we use fixed effects for each firm. In line with Sprenger et al. (2014b), we find significant impact of sentiment in tweets on stock returns and the magnitude of the effect is large. A 1% increase in positiveness of Twitter messages leads to a 0.0239% increase in daily returns. Wysocki (1998) document a significant relation between message volume of stock message boards and stock returns. On the contrary, we find no significant relationship between Twitter messages volume and stock returns.[6]

[6] We also conduct Fama-MacBeth regressions which yield qualitatively similar results. (available upon request)

Similarly, Table 3 shows fixed effects panel regressions of volatility, volume and bid-ask spreads as dependent variable and the four tweet features as independent variables. There are strongly significant impacts of tweet indicators on high-low range and trading volume. The coefficient of message volume is positive and statistically significant in explaining variation in volatility and trading volume. The bots indicator of tweet feature can also significantly affect ex-post volatility and trading volume. A 1% increase in bots results in a 0.15% increase in volatility, and a 0.04% increase in trading volume. The magnitude of the impact is larger than that of FTSE 100. And the impact is also of economic significance due to the fact that bot-tweets could easily be increased or tuned by Twitter farms or bot farms. Contrary to Sprenger et al. (2014b), we find statistically significant relations of the agreement measure with volatility and trading volume.

5.2. Robustness check

There is possibility of endogeneity that higher returns could cause higher degree of positive information dissemination in Twitter. In order to mitigate the potential endogeneity issues, we do robustness checks by separating the Twitter postings into two groups based on the opening time of London Stock Exchange: from 4:30 pm yesterday to 8:00am today as pre-trading, and between 8:00am and 4:30pm today as trading. We then run contemporaneous regressions by using the tweet features collected in the pre-trading hours, and obtain similar results as reported in Table 4 and 5. There is no significant impact of tweet indicators on stock returns, but we find statistically significant relations between tweet features (positiveness, message volume, agreement and bots) and market features (volatility and trading volume).

We then perform further robustness checks by investigating the lagged relationships between tweet and market features, and one day lagged tweet features are regressed on market features. We obtain similar results as given in Table A4 and A5 in Appendix A. There is no statistically significant association between lagged positiveness and stock returns. Contrary to Antweiler and Frank (2004), we find little statistically significant relation between message volume and stock returns. Sprenger et al. (2014b) add that the reason for this is because investors are more thoughtful when assessing the information content of Twitter messages compared to online stock message boards. Enikolopov et. al (2016) also argue that institutional investors are less likely to be influenced by Twitter, and non-institutional investors, who are more likely to be affected by tweet postings, may only change trading volume but not returns.

Moreover, we find statistically significant associations between one-day lagged positiveness, message volume, bots and volatility, trading volume. We also use an alternative measure of bots, which is defined as the proportion of bot-tweets over the total number of tweets on a day, as opposed to the natural logarithm of the volume of bot-tweets employed previously, and obtain similar results as shown in Table A5 to A8 in Appendix A.

5.3. Event study

We also use event study to identify abnormal increases in tweets and bot-tweets volume, and examine the impacts of event on stock market features. An event or abnormal increase in tweets or bot-tweets volume satisfies the following three conditions: 1) the absolute number of tweets or bot-tweets is in the top 5% of each company's empirical distribution of daily tweets; 2) the relative increase in tweets or bot-tweets volume is larger than 100%; 3) the absolute number of increase in tweets or bot-tweets volume is greater than 500 or 100. Figure 2 depicts the number of events when there are abnormal increases in tweets containing a FTSE 100 firm name. There are about one to two events per day on 55 sampled firms during the majority of our sample period.

Panel A of Table 6 reports reactions in returns following different types of events. There is no significant reaction on the event days (0) when there are abnormal increases in the volume of positive tweets. In contrast, stock prices decrease significantly in response to abnormal increases in negative tweets volume. All reactions are reported in percentages, hence the magnitude of reduction is 0.67%. However, we find significant positive returns during the following week. (1, 5) The stock prices recover to almost the same level before the event in the following week. Interestingly, abnormal increases of retweets volume causes this bounce-back pattern of stock prices whereas original tweets do not. Panel B of Table 6 reports equivalent event study results while returns are mean-adjusted. The results are consistent with those in Panel A and confirm the bounce-back pattern of stock prices associated with abnormal increases in the volume of negative tweets, especially retweets. Panel C of Table 6 confirms the same bounce-back pattern of stock prices in response to abnormal increases in negative retweets volume when we employ market-model abnormal return.

Panel D of Table 6 presents responses of ex-post volatility measure, to abnormal increases in tweets volume. There are strongly significant and positive reactions in all time windows. This implies that abnormal increases in the volume of both positive and negative tweets, original tweets and retweets could deepen the ex-post disagreements among market participants. Panel E of Table 6 shows reactions of normalized trading volume to abnormal increases in tweets volume. There are statistically significant and economically meaningful reactions of trading volume during all time windows after abnormal increases in tweets volume regardless of the type of tweets: original tweets or retweets, bot-tweets or human-tweets, and the magnitude of the reactions is huge.[7] Panel F of Table 6 reports responses of bid-ask spreads, which is an inversed measure of liquidity, to abnormal increases in tweets volume. Again, we find strongly significant responses of bid-ask spreads.[8]

Overall, evidence from the event study confirms the phenomena that information embedded in tweets imposes strong impact on market participants' disagreements, i.e. ex-post volatility, trading volume and liquidity, while there is bounce-back pattern of price reactions only in

[7] One week after abnormal increases in the volume of all tweets, trading volume increases over 18.8%+31%=49.8%.

[8] Over the week after the event, bid-ask spreads decrease by 10.7+45.4=56.1 basis points.

response to negative retweets. Again, the findings support the argument that tweets are more likely to affect non-institutional investors.

5.4. Intraday analysis

The real-time nature of Twitter indicates that we shall also examine intraday stock prices. Since we aggregate all daily tweet postings to obtain the sentiment of tweets, sentiment could thus influence the stock market during the day. And the tweet postings after the market closes are assigned to the following day, consequently there is a time lag before these Twitter postings can influence the stock market.

Table 7 presents regression results based on 5-minute data from February to July 2017. There is no significant impact of tweets and bot-tweets on returns, while significant coefficients of tweets and bot-tweets indicators in explaining variation in ex-post volatility and trading volume indicate that information embedded in tweets disseminates in the stock market mainly through trading volume and disagreements among market participants.

Table 8 shows results by regressing one-interval lagged tweet features on market features based on 5-minute data. In contrast to results in Table 7, we find significant causality from tweets and bot-tweets to stock returns. While volatility and trading volume are still affected by information in tweets and bot-tweets.

Table 9 reports regression results by using one to six intervals lagged independent variables. Notably, coefficients of lagged bot on returns are significant and negative, but stop at lag six and become positive for log return as dependent variable. This suggests that lagged increases in bot-tweets lead to significant reductions of stock returns and the effects diminish after 15 or 25 minutes if we employ mean-adjusted or market-model return as the dependent variable. These findings are of significant importance as they could not be detected when we use daily data. They also justify insignificant impacts of tweets and bot-tweets on stock prices discussed before. Lastly, we perform contemporaneous regressions and lagged variables regressions using the tweet features collected in pre-trading hours. We obtain similar results as shown in Table A9 and A10 in Appendix A.

6. Conclusions

Social media has become a popular platform for information sharing and acquisitions. However, its convenience and popularity also come with threats. Recent literature and events have raised attentions on the uses and abuses of social media for cyber interferences in Western democracies. Social media bots/farms could be weaponized during constitutional referendums, elections and swinging political opinions (e.g. U.S Intelligence Committee, 2018). Scientific evidence of bots' impacts on actual outcomes, however, is limited given the fact that political beliefs are hard to measure at a reasonable frequency. This paper employs FTSE 100 composites as an alternative measure of actual outcomes. Specifically, we investigate whether volume and sentiments of tweets/bot-tweets trigger abnormal changes in stock indicators using a sample of 55 companies in the FTSE 100 composites during the period from August 2015 to July 2017.

Based on the daily frequency, we find insignificant associations between tweets/bot-tweets and stock returns. There are statistically significant relations between the sentiments of tweets and bot-tweets, and stock volatility and trading volume. This indicates that information embedded in social media can help to forecast volatility and trading volume. We also use 5-minute stock prices to perform intraday analysis, and find a positive (negative) impact of tweets (bot-tweets) on stock returns. The negative impact of bot-tweets vanishes and reverses within 30 minutes. The results for volatility and trading volume are consistent with the daily data analysis. We also adopt event study which reveals a bounce-back pattern of price reactions in response to negative retweets. Abnormal increases in tweets/bot-tweets have significant effects on stock volatility, trading volume and liquidity. Our results are robust against numbers of alternative specifications. Tweets are collected before trading hours, i.e. before 8:00 am, to mitigate endogeneity problems. Lagged features of tweets/bot-tweets are employed to further clarify the economic forces at work.

In line with existing literature that focuses on message volume and sentiment (Wysocki, 1998; Sprenger et al., 2014), we also use event study to identify abnormal increases in tweets and bot-tweets volume and find that the information embedded in tweets has strong impact on market participants' disagreements, i.e. ex-post volatility, trading volume and liquidity, and there is a bounce-back pattern of price reactions in response to increases in the volume of negative retweets.

Taken together, our findings reveal evidence of social media bots' influence on real outcomes. Particularly, the uses and abuses of social media bots provoke instability i.e. intensify public heterogeneous beliefs and opinions polarization. Small investors rather than institutional investors are more likely to be influenced by information spread in social media. Moreover, policy makers need to enforce the transparency of advertisements published on social media and establish proper code of practice for social media, to prevent the potential spread of fake information. Policy makers should have mechanisms to monitor the use of social bots as social bots could influence the public in various ways. In addition, there are needs for regulatory measures to raise social media literacy and attentions on the uses and abuses of social media bots for future developments on safeguarding public welfare, especially small investors' interests. Lastly, our study suggests potential benefits of social media data availability for future research.

This study also has limitations and there are some potential areas for future work. Firstly, there should be caveat in interpretations of the results. Sprenger et al. (2014b) argue that large number of observations often result in statistical significance despite large cross-sectional differences in financial measures such as returns. Hence, we may not expect the significant relation between message volume and trading volume to hold for every company. Secondly, since our findings support the argument that tweets are more likely to affect non-institutional investors, we could use the number of trades of different size categories to separate between institutional and non-institutional investors. We leave these questions for future research.

Table 1: Descriptive statistics

This table reports summary statistics of the full sample and high versus low bot activities subsamples in Panel A and B respectively. The sample covers 55 of FTSE 100 composites from 1st August 2015 to 31st July 2017. High-bot (low-bot) denotes at least (at most) 50% of the total tweets. *, **, *** respectively, denotes 5%, 1%, 0.1% significance based on unequal variances t-tests.

	Mean	Std. Dev.	No. of Obs.
P/E ratio	25.1126	42.1919	24,217
Earnings per share	68.0268	73.2717	24,640
Book value	6.7853	10.1250	24,640
ROE (%)	14.0117	27.9366	24,192
Leverage (Debt/Equity, %)	121.3340	218.5305	24,640
Return (%)	0.0149	2.0000	24,640
Log(volume)	8.0179	2.1124	24,640
High-low range (%)	1.7009	1.2715	23,760
Bid-ask spreads (basis points)	5.4715	4.0657	23,760
No. of tweets	2015.1796	4904.0930	24,640
Positiveness	1.1029	1.0673	24,640
Agreement	0.2080	0.2003	24,640
No. of tweets generated by bots	129.4942	452.2963	24,640
Positiveness from bot-tweets	0.6268	1.1504	24,640
Agreement from bot-tweets	0.1705	0.2112	24,640

Table 2: Regressions of returns by tweets

This table reports fixed-effects regressions of stock returns by tweet characteristics. Dependent variables are (log)return, mean-adjusted return, market-model return. Main independent variables are tweet characters based on tweets collected from 4:30pm day -1 to 4:29pm day 0. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (tweets generated by automatic algorithm accounts). Bots × Positiveness (Bots × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

Variable	Return (1)	Return (2)	Mean-adj. return (3)	Mean-adj. return (4)	Mrk-model return (5)	Mrk-model return (6)
Positiveness	0.0239*	0.0532*	0.0207*	0.0471	0.0251*	0.0435
	(2.49)	(2.02)	(2.15)	(1.80)	(2.17)	(1.39)
Message	-0.0324	-0.0387	-0.0174	-0.0311	-0.0308	-0.0392
	(-1.33)	(-1.48)	(-0.71)	(-1.19)	(-1.06)	(-1.25)
Agreement	-0.0088	-0.0137	-0.0107	-0.0247	-0.0102	-0.0073
	(-1.00)	(-0.41)	(-1.21)	(-0.75)	(-0.96)	(-0.18)
FTSE 100 return	0.5495***	0.5493***	0.5491***	0.5489***	0.0333***	0.0331***
	(69.12)	(69.20)	(69.25)	(69.34)	(3.56)	(3.55)
Bots		0.0010		0.0154		0.0051
		(0.07)		(1.05)		(0.29)
Bots × Positiveness		0.0348		0.0301		0.0227
		(1.49)		(1.29)		(0.82)
Bots × Agreement		-0.0065		-0.0169		0.0018
		(-0.21)		(-0.54)		(0.05)
Observations	24,089	24,089	24,089	24,089	24,089	24,089
R^2	0.304	0.304	0.302	0.303	0.002	0.002

Table 3: Regressions of volatility, trading volume, bid-ask spreads by tweets

This table reports regression of stock volatility, trading volume, bid-ask spreads by tweets. Dependent variables in columns (1), (2) are volatility measure i.e. Parkinson (1980) intraday high-low range. Dependent variables in columns (3)-(4) and (5)-(6) are trading volume and bid-ask spreads respectively. Main independent variables are based on tweets collected from 4:30pm day -1 to 4:29pm day 0. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (bot-tweets). Bots × Positiveness (Bots × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Volatility	(2) Volatility	(3) Volume	(4) Volume	(5) Bid-Ask	(6) Bid-Ask
Positiveness	-0.0387***	-0.0533	-0.0027	-0.0131	-0.0143	0.0229
	(-3.37)	(-1.81)	(-0.96)	(-1.61)	(-1.70)	(1.07)
Message	0.3841***	0.2941***	0.0739***	0.0489***	-0.0162	-0.0273
	(14.50)	(9.92)	(12.32)	(7.24)	(-0.91)	(-1.41)
Agreement	-0.0158	-0.0916*	-0.0166***	-0.0255*	-0.0005	0.0031
	(-1.59)	(-2.57)	(-5.38)	(-2.54)	(-0.05)	(0.11)
FTSE 100 return	-0.0871***	-0.0883***	-0.0130***	-0.0133***	0.0090	0.0087
	(-8.72)	(-8.86)	(-6.38)	(-6.57)	(1.82)	(1.76)
Bots		0.1578***		0.0419***		0.0026
		(9.44)		(8.00)		(0.17)
Bots × Positiveness		-0.0273		-0.0137		0.0457*
		(-1.02)		(-1.66)		(2.04)
Bots × Agreement		-0.0884*		-0.0114		0.0017
		(-2.56)		(-1.11)		(0.06)
Observations	23,242	23,242	23,242	23,242	23,242	23,242
R^2	0.239	0.243	0.915	0.915	0.227	0.227

Table 4: Pre-trading regressions for returns

This table reports fixed-effects regressions of stock returns by tweets. Dependent variables are (log)return, mean-adjusted return, market-model return. Main independent variables are based on tweets collected from 4:30pm day -1 to 8:00am day 0. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (bot-tweets). Bots × Positiveness (Bots × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Return	(2) Return	(3) Mean-adj. return	(4) Mean-adj. return	(5) Mrk-model return	(6) Mrk-model return
Positiveness	0.0141	0.0147	0.0104	0.0067	0.0142	0.0050
	(1.55)	(0.61)	(1.15)	(0.28)	(1.30)	(0.17)
Message	-0.0326	-0.0417	-0.0192	-0.0340	-0.0338	-0.0444
	(-1.56)	(-1.74)	(-0.92)	(-1.42)	(-1.35)	(-1.55)
Agreement	-0.0081	-0.0161	-0.0099	-0.0257	-0.0105	-0.0162
	(-0.91)	(-0.49)	(-1.10)	(-0.77)	(-0.97)	(-0.41)
FTSE 100 return	0.5498***	0.5497***	0.5493***	0.5491***	0.0338***	0.0337***
	(67.79)	(67.87)	(67.91)	(67.99)	(3.54)	(3.53)
Bots		0.0158		0.0278		0.0199
		(0.99)		(1.74)		(1.04)
Bots × Positiveness		0.0001		-0.0056		-0.0109
		(0.00)		(-0.23)		(-0.37)
Bots × Agreement		-0.0092		-0.0180		-0.0069
		(-0.28)		(-0.55)		(-0.17)
Observations	23,592	23,592	23,592	23,592	23,592	23,592
R^2	0.304	0.304	0.302	0.303	0.002	0.002

Table 5: Pre-trading regressions for volatility, trading volume, bid-ask spreads

This table reports regression of stock volatility, trading volume, bid-ask spreads by tweets. Dependent variables in columns (1), (2) are volatility measure i.e. Parkinson (1980) intraday high-low range. Dependent variables in columns (3) - (4) and (5) - (6) are trading volume and bid-ask spreads, respectively. Main independent variables are based on tweets collected from 4:30pm day -1 to 8:00am day 0. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (bot-tweets). Bots × Positiveness (Bots × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Volatility	(2) Volatility	(3) Volume	(4) Volume	(5) Bid-Ask	(6) Bid-Ask
Positiveness	-0.0297**	-0.0579*	-0.0029	-0.0196*	-0.0175*	0.0473
	(-2.67)	(-1.99)	(-1.09)	(-2.27)	(-2.16)	(1.93)
Message	0.2905***	0.2212***	0.0514***	0.0363***	-0.0178	-0.0292
	(12.21)	(7.66)	(8.89)	(5.57)	(-1.02)	(-1.50)
Agreement	-0.0186	-0.0744*	-0.0151***	-0.0161	0.0010	-0.0186
	(-1.82)	(-2.09)	(-4.96)	(-1.44)	(0.09)	(-0.51)
FTSE 100 return	-0.0914***	-0.0922***	-0.0139***	-0.0141***	0.0082	0.0080
	(-8.91)	(-9.00)	(-6.66)	(-6.76)	(1.64)	(1.59)
Bots		0.1274***		0.0272***		0.0048
		(6.67)		(5.21)		(0.31)
Bots × Positiveness		-0.0368		-0.0190*		0.0725**
		(-1.18)		(-2.19)		(2.82)
Bots × Agreement		-0.0646		-0.0021		-0.0224
		(-1.79)		(-0.18)		(-0.55)
Observations	22,761	22,761	22,761	22,761	22,761	22,761
R^2	0.228	0.231	0.914	0.914	0.226	0.227

Table 6: Event study – Market responses following abnormal surges in tweet activities

This table reports average cumulative (abnormal) returns, average cumulative changes in volatility, trading volume, bid-ask spreads in responses to abnormal increases in numbers of tweets. An abnormal increase in tweets satisfies all the following three conditions: (i) in the top 5% of the empirical distribution of daily changes in each firm; (ii) relative change is larger than 100%; (iii) absolute change is larger than 500 (100 for bot activities). [0], [1], [1,5] report average cumulative changes in percentage points. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

Time windows	All	Positive	Negative	Bots	Bot pos.	Bot neg.
Panel A: Response of returns						
[0]	-0.148	-0.228	-0.670**	-0.282	-0.121	-0.269
[1]	0.025	0.143	-0.069	-0.018	0.159	-0.068
[1,5]	0.479**	0.204	0.663*	0.724**	0.347	0.995**
Obs.	751	426	306	361	153	77
Panel B: Response of mean-adjusted returns						
[0]	-0.134	-0.212	-0.653**	-0.261	-0.125	-0.264
[1]	0.040	0.161	-0.050	0.005	0.157	-0.059
[1,5]	0.538***	0.274	0.739**	0.811***	0.337	1.029**
Obs.	751	426	306	361	153	77
Panel C: Response of market-model abnormal returns						
[0]	-0.033	-0.073	-0.530*	-0.122	-0.070	-0.317
[1]	0.008	0.104	-0.061	-0.013	0.155	0.039
[1,5]	0.320**	0.092	0.311	0.463**	-0.039	0.852**
Obs.	751	426	306	361	153	77
Panel D: Response of volatility						
[0]	0.616***	0.554***	0.690***	0.436***	0.306*	0.078
[1]	0.273***	0.220***	0.376***	0.201**	0.135	0.099
[1,5]	0.524***	0.601***	0.541**	0.612***	0.253	-0.073
Obs.	718	417	293	352	149	72
Panel E: Response of (normalized) trading volume						
[0]	18.81***	6.05***	15.11***	9.28***	3.77	2.64
[1]	13.27***	10.13***	16.69***	9.44***	4.73	4.83
[1,5]	31.04***	25.67***	41.05***	22.28**	11.44	5.49
Obs.	718	417	293	352	149	72
Panel F: Response of bid-ask spreads						
[0]	-10.75***	-8.46***	-9.18***	-10.64***	-5.87	-7.41*
[1]	-10.84***	-10.40***	-3.51	-4.16	3.37	-7.09
[1,5]	-45.43***	-37.03***	-32.04***	-29.62***	-34.37***	-12.15
Obs.	718	417	293	352	149	72

Table 7: Intraday regressions

This table reports fixed-effects regressions of stock returns, high-low range, trading volume by tweets based on 5-minute data from February to July 2017. Dependent variables are (log)return, mean-adjusted return, market-model return, volatility, (normalized) trading volume. Main independent variables are based on tweets collected during 5-minute intervals. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (bot-tweets). Bots × Positiveness (Bots × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Return	(2) Return	(3) Mean-adj. return	(4) Mean-adj. return	(5) Mrk-model return	(6) Mrk-model return	(7) Volatility	(8) Volatility	(9) Volume	(10) Volume
Positiveness	0.0005	-0.0041	-0.0007	-0.0006	-0.0012	0.0015	-0.0008*	0.0810***	0.0089**	0.2372***
	(0.82)	(-0.41)	(-1.19)	(-0.06)	(-1.95)	(0.14)	(-2.34)	(16.75)	(2.59)	(9.72)
Message	-0.0009	-0.0009	-0.0010	-0.0008	-0.0004	-0.0000	0.0221***	0.0152***	0.0772***	0.0692***
	(-1.12)	(-1.38)	(-1.17)	(-1.17)	(-0.42)	(-0.03)	(53.33)	(43.98)	(32.58)	(25.00)
Agreement	-0.0006	0.0153	0.0014	0.0200	0.0022*	0.0233	0.0038***	-0.2311***	-0.0135*	-0.6156***
	(-0.52)	(0.95)	(1.32)	(1.23)	(2.02)	(1.42)	(6.23)	(-27.35)	(-2.02)	(-10.28)
FTSE 100 return	83.3407***	83.3623***	83.3524***	83.3770***	-22.4987***	-22.4682***	3.8962***	3.2208***	-6.5347*	-7.8346**
	(62.19)	(62.31)	(61.98)	(62.10)	(-15.53)	(-15.53)	(5.53)	(4.71)	(-2.26)	(-2.73)
Bots		-0.0004		-0.0015		-0.0023		0.0267***		0.0375***
		(-0.24)		(-0.82)		(-1.25)		(31.14)		(6.68)
Bots × Positiveness		-0.0007		0.0001		0.0005		0.0122***		0.0344***
		(-0.46)		(0.04)		(0.29)		(16.89)		(9.31)
Bots × Agreement		0.0024		0.0027		0.0031		-0.0347***		-0.0897***
		(0.99)		(1.13)		(1.26)		(-27.69)		(-10.18)
N	574,635	574,635	574,635	574,635	574,635	574,635	574,635	574,635	574,634	574,634
R^2	0.117	0.117	0.328	0.328	0.237	0.237	0.110	0.124	0.603	0.603

Table 8: Regressions for 5-minute lagged tweets

This table reports fixed-effects regressions of stock returns, high-low range, trading volume by tweets based on 5-minute data from February to July 2017. Dependent variables are (log)return, mean-adjusted return, market-model return, volatility, (normalized) trading volume. Main independent variables are one-lagged features of tweets collected during 5-minute intervals. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (bot-tweets). Bots × Positiveness (Bots × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Return	Return	Mean-adj. return	Mean-adj. return	Mrk-model return	Mrk-model return	Volatility	Volatility	Volume	Volume
Lagged Positiveness	-0.0004	-0.0016	-0.0016**	0.0019	-0.0021***	-0.0007	-0.0004	0.0393***	0.0037	0.1359***
	(-0.77)	(-0.36)	(-3.29)	(0.43)	(-4.51)	(-0.16)	(-1.44)	(13.64)	(1.08)	(5.73)
Lagged Message	0.0012**	0.0009*	0.0011**	0.0010*	-0.0003	0.0003	0.0126***	0.0091***	0.0430***	0.0421***
	(2.72)	(2.10)	(2.59)	(2.37)	(-0.76)	(0.75)	(44.99)	(34.30)	(18.57)	(15.45)
Lagged Agreement	0.0004	-0.0294**	0.0024*	-0.0249*	0.0030**	-0.0046	0.0014*	-0.1204***	-0.0161*	-0.4919***
	(0.44)	(-3.01)	(2.50)	(-2.50)	(3.20)	(-0.47)	(2.41)	(-18.97)	(-2.44)	(-8.32)
Lagged FTSE100return	1.9786**	1.9386**	2.0219**	1.9849**	-1.1052	-1.0846	1.1609**	0.8112	0.2516	-0.4478
	(3.16)	(3.10)	(3.21)	(3.15)	(-1.85)	(-1.82)	(2.60)	(1.86)	(0.10)	(-0.18)
Lagged Bots		0.0025*		0.0015		-0.0016		0.0140***		0.0166**
		(2.57)		(1.50)		(-1.71)		(22.57)		(2.97)
Lagged (Bots × Positiveness)		-0.0002		0.0005		0.0002		0.0059***		0.0199***
		(-0.35)		(0.75)		(0.35)		(13.61)		(5.52)
Lagged (Bots ×Agreement)		-0.0044**		-0.0041**		-0.0012		-0.0180***		-0.0710***
		(-3.02)		(-2.73)		(-0.80)		(-19.06)		(-8.17)
N	574,416	574,416	574,416	574,416	574,412	574,412	574,416	574,416	574,415	574,415
R^2	0.000	0.000	0.246	0.246	0.232	0.232	0.088	0.092	0.602	0.602

Table 9: Regressions for 30-minute lagged tweets

This table reports fixed-effects regressions of stock returns, high-low range, trading volume by tweets based on data from February to July 2017. Dependent variables are (log)return, mean-adjusted return, market-model return, volatility, (normalized) trading volume. Main independent variables are six lagged-bot-activities based on tweets collected during 5-minute intervals. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (bot-tweets). Bots × Positiveness (Bots × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Return	(2) Return	(3) Mean-adj. return	(4) Mean-adj. return	(5) Mrk-model return	(6) Mrk-model return	(7) Volatility	(8) Volatility	(9) Volume	(10) Volume
Lagged Positiveness	-0.0008	-0.0023	-0.0014*	-0.0002	-0.0017**	-0.0026	-0.0005	0.0353***	-0.0005	0.1113***
	(-1.55)	(-0.50)	(-2.56)	(-0.05)	(-3.14)	(-0.58)	(-1.63)	(11.41)	(-0.14)	(4.60)
Lagged Message	0.0019***	0.0015**	0.0019***	0.0016**	-0.0003	0.0000	0.0106***	0.0081***	0.0265***	0.0258***
	(3.81)	(3.01)	(3.80)	(3.18)	(-0.69)	(0.04)	(31.10)	(25.24)	(10.48)	(8.42)
Lagged Agreement	0.0008	-0.0299**	0.0019	-0.0264**	0.0022*	-0.0095	0.0013*	-0.1077***	-0.0049	-0.4158***
	(0.81)	(-3.04)	(1.89)	(-2.62)	(2.23)	(-0.96)	(2.21)	(-16.53)	(-0.72)	(-6.92)
Lagged 1 Bots		0.0024*		0.0017		-0.0008		0.0123***		0.0220***
		(2.39)		(1.65)		(-0.81)		(18.63)		(3.80)
Lagged 2 Bots		-0.0030***		-0.0037***		-0.0032***		0.0085***		0.0117*
		(-3.48)		(-4.15)		(-3.71)		(15.42)		(2.02)
Lagged 3 Bots		-0.0016		-0.0022**		-0.0024**		0.0074***		0.0298***
		(-1.90)		(-2.60)		(-2.93)		(14.34)		(5.15)
Lagged 4 Bots		0.0010		0.0002		-0.0005		0.0046***		0.0097
		(1.22)		(0.30)		(-0.58)		(9.44)		(1.66)
Lagged 5 Bots		-0.0009		-0.0016*		-0.0014		0.0037***		0.0107
		(-1.15)		(-2.02)		(-1.82)		(7.73)		(1.85)
Lagged 6 Bots		0.0022**		0.0014		0.0005		0.0033***		0.0097
		(2.99)		(1.85)		(0.67)		(7.06)		(1.69)
N	573,304	573,304	573,304	573,304	573,300	573,300	573,304	573,304	573,303	573,303
R^2	0.000	0.001	0.247	0.247	0.232	0.232	0.089	0.098	0.602	0.603

Figure 1: Weekly number of tweets

This figure describes weekly number of different types of tweet messages on the 55 sampled firms from FTSE 100 composites from August 2015 to July 2017. There were two weeks of tweeting activities i.e. week commencing Oct 19, 2015 and, Sep 12, 2016.

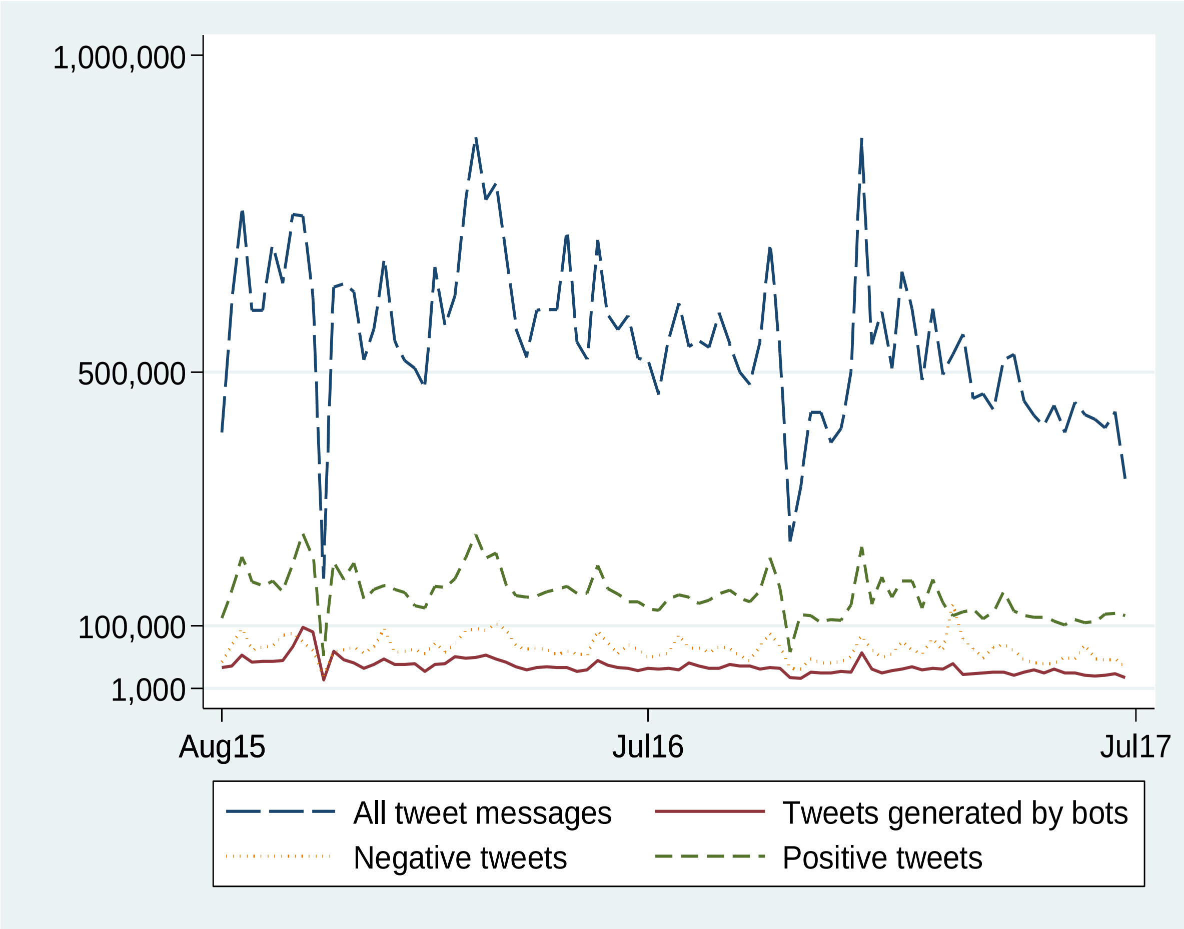
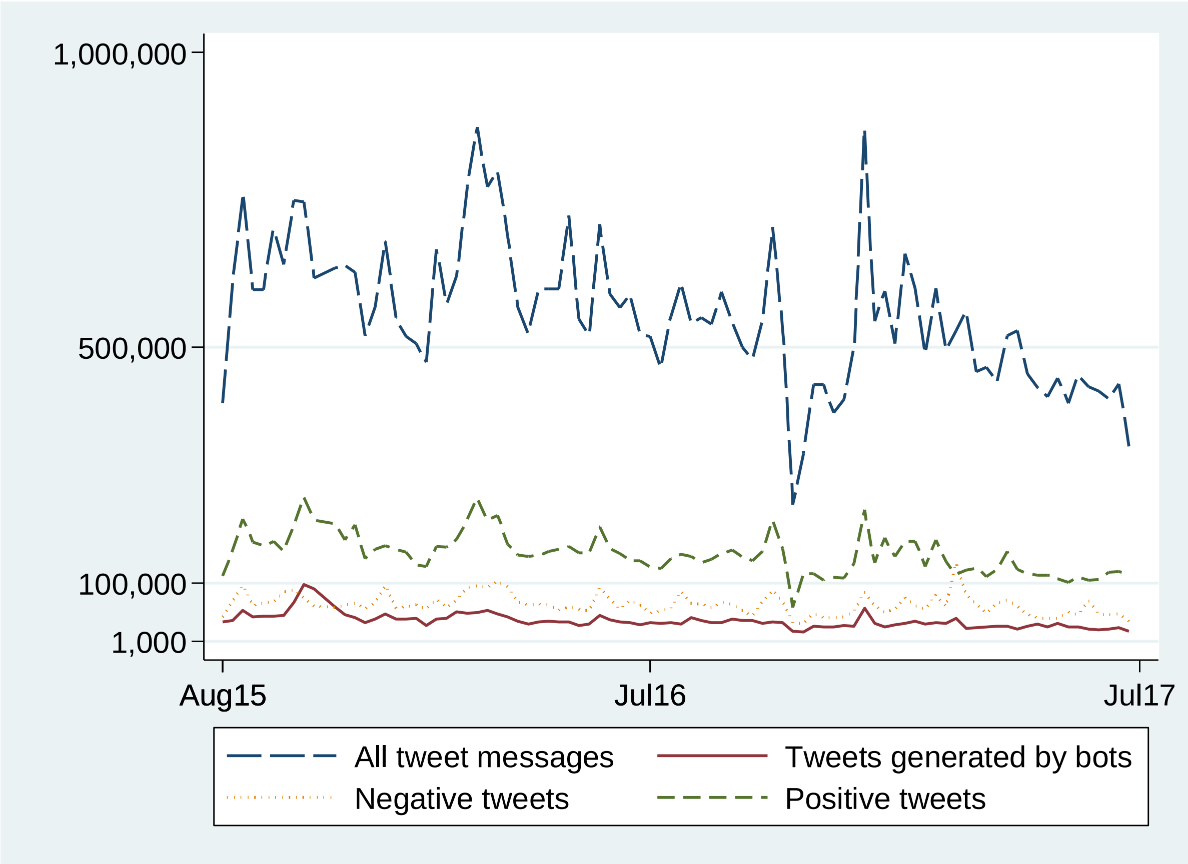


Figure 1: Weekly number of tweets[9]

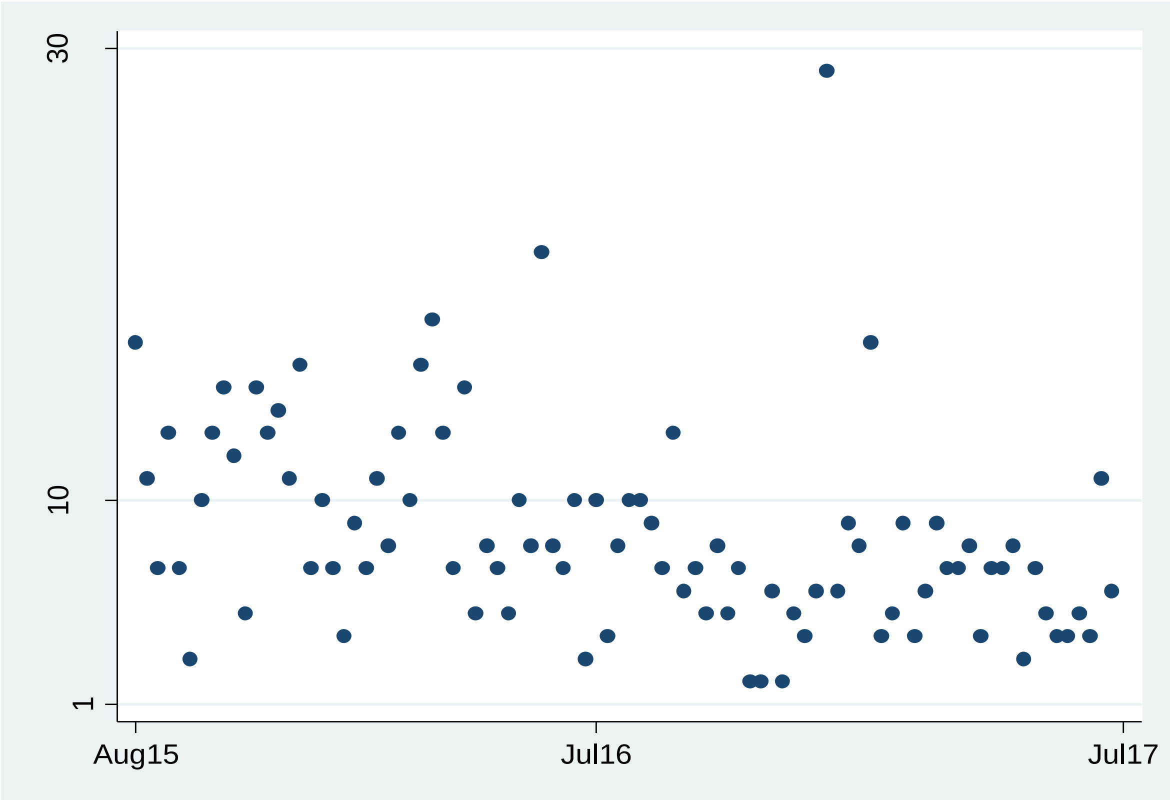
This figure describes weekly number of different types of tweet messages on the 55 sampled firms from FTSE 100 composites from August 2015 to July 2017. During week commencing Sep 12, 2016 there are less tweet activities i.e. number of tweet messages was lower than 300,000.



[9] In this version of figure 1, I drop a week in Oct15 when the computer was down and there were two days without tweets during the week. Please choose between the two figures

Figure 2: Abnormal increases in tweets

This figure depicts number of events per week. An event is defined as a day when there are abnormal increases in tweets containing a FTSE 100 firm name which satisfies three conditions: (i) in top 5% of empirical distribution of daily changes in tweets for each sampled firm; (ii) relative increase is larger than 100%; (iii) absolute increase is larger than 500. There were 29 events happened during week commencing Jan 23, 2017.



References

Antweiler, W. and Frank, M.Z., 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance*, 59, 1259-1294.

Barber, B. and Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, 21, 785-818.

Blankespoor, E., Miller, G. and White, H., 2014. The role of dissemination in market liquidity: evidence from firm's use of Twitter. *Accounting Review*, 89, 79-112.

Bollen, J., Mao, H. and Zeng, X−J., 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2, 1-8.

Bonaparte, Y. and Kumar, A. (2013) Political activism, information costs, and stock market participation. *Journal of Financial Economics*, 107, 760–786.

Brown, G.W., 1999. Volatility, sentiment, and noise traders. *Financial Analysts Journal*, 55, 82-90.

Bushee, B., Core, J.E., Guay, W. and Hamm, S.J., 2010. The role of the business press as an information intermediary. *Journal of Accounting Research*, 48, 1-19.

Chen, H., De, P., Hu, Y.J. and Hwang, B.H., 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, 27, 1367-1403.

Da, Z., Engelberg, J. and Gao, P., 2011. In search of attention. *Journal of Finance*, 66, 1461-1499.

Danthine, J−P. and Moresi, S., 1993. Volatility, information, and noise trading. *European Economic Review*, 37, 961-982.

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J., 1990. Noise trader risk in financial markets. *Journal of Political Economy*, 98, 703-738.

Dewally, M., 2003. Internet investment advice: Investing with a rock of salt. *Financial Analysts Journal*, 59, 65-77.

Demarzo, P.M, Vayanos, D., Zwiebel, J., 2003. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics,* 118, 909-968

Dougal, C., Engelberg, J., Garcia, D. and Parsons, C.A., 2012. Journalists and the stock market. *Review of Financial Studies*, 25, 639-679.

Engelberg, J.E. and Parsons, C.A., 2011. The causal impact of media in financial markets. *Journal of Finance*, 66, 67-97.

Enikolopov, R., Petrova, M. and Sonin, K., 2016. Social media and corruption. *Working paper*, University Pompeu Fabra and University of Chicago.

Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A., 2016. The rise of social bots. *Communications of the ACM*, 59, 96-104.

Forbes, 2017. Can 'fake news' impact the stock market? Available at https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/#6cc91dc72fac (accessed March 26, 2017).

Gorodnichenko, Y., Pham, T. and Talavera, O., 2017. Social media, sentiment and public opinions: Evidence from Brexit and US Election. *Working paper*, University of Californian Berkeley and Swansea University.

Gross-Klussmann A. and Hautsch N., 2011. When machines read the news: using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, 18, 321-340.

Hirshleifer, D. and Teoh, S., 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36, 337-386.

Hong, H., Kubik, J. and Stein, J.C., 2005. The neighbor’s portfolio: word-of-mouth effects in the holdings and trades of money managers. *Journal of Finance*, 60, 2801-2824.

Howard, N.P, Bolsover, G., Kollanyi, B., Bradshaw, S., and Neudert, L-M., 2017. Junk news and bots during the U.S. Election: What were Michigan voters sharing over Twitter? *Working* *paper*, Oxford University.

Kerl, A.G. and Walter, A., 2007. Market responses to buy recommendations issued by personal finance magazines: effects of information, price-pressure, and company characteristics. *Review of Finance*, 11, 117-141.

Koski, J.L., Rice, E.M. and Tarhouni, A., 2004. Noise trading and volatility: Evidence from day trading and message boards. *Working paper,* University of Washington.

Ng, L. and Wu, F., 2006. Peer effects in the trading decisions of individual investors. *Financial Management*, 39, 807-831.

Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 53, 61-65.

Pritamani, M. and Singal, V., 2001. Return predictability following large price changes and information releases. *Journal of Banking and Finance*, 25, 631-651.

Ranco, G. Aleksovski, D., Caldarelli, G., Grcar, M. and Mozetic, I., 2015. The effects of Twitter sentiment on stock price returns. *Plos One*, 10, 1-21.

Ryan, P. and Taffler, R., 2004. Are economically significant stock returns and trading volumes driven by firm-specific news releases? *Journal of Business Finance and Accounting*, 31, 49-81.

Sims, C.A., 2003. Implications of rational inattention. *Journal of Monetary Economics*, 50, 665-690.

Sprenger, T.O., Sandner, P.G., Tumasjan, A. and Welpe, I.M., 2014a. News or noise? The stock market reaction to different types of company-specific news events. *Journal of Business Finance and Accounting*, 41, 791-830.

Sprenger, T.O., Tumasjan, A., Sandner, P.G. and Welpe, I.M., 2014b. Tweets and trades: the information content of stock microblogs. *European Financial Management*, 20, 926-957.

Solomon, D., Soltes, E. and Sosyura, D., 2014. Winners in the spotlight: media coverage of fund holdings as a driver of flows. *Journal of Financial Economics*, 113, 53-72.

Tetlock, P.C., 2007. Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance*, 62, 139-168

Tetlock, P.C., 2010. Does public financial news resolve asymmetric information? *Review of Financial Studies*, 23, 3520–3557.

Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S., 2008. More than words: quantifying language to measure firms' fundamentals. *Journal of Finance*, 63, 1437–1467.

Tumarkin, R. and Whitelaw, R.F., 2001. News or noise? Internet postings and stock prices. *Financial Analysts Journal*, 57, 41-51.

U.S. Intelligence Committee, 2018. Open hearing: Worldwide threats hearing. Available at https://www.intelligence.senate.gov/hearings/open-hearing-worldwide-threats-hearing-1 (accessed February 13, 2018).

Wysocki, P., 1998. Cheap talk on the web: The determinants of postings on stock message boards. *Working paper*, University of Michigan.

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

Appendix A

Table A1: Sampled companies

This table lists names of the sampled companies. The sampled period is from July 2015 to August 2017.

	Firm name			Firm name
1	3I		31	Lloyds
2	Anglo American		32	London Stock Exchange
3	Antofagasta		33	Marks and Spencer
4	Astrazeneca		34	Mondi
5	Babcock		35	Morrison
6	BAE Systems		36	National Grid
7	Barclays		37	Old Mutual
8	BHP Billiton		38	Pearson
9	British American Tobacco		39	Persimmon
10	British Land		40	Prudential
11	British Petroleum		41	Reckitt Benckiser
12	BT Group		42	Rio Tinto
13	Bunzl		43	Rolls-Royce
14	Burberry		44	Royal Bank of Scotland
15	Carnival		45	Royal Dutch Shell
16	Centrica		46	Royal Mail
17	Coca-Cola		47	Sainsbury
18	Compass Group		48	Schroders
19	Diageo		49	Severn Trent
20	Direct Line		50	Standard Chartered
21	Easyjet		51	Taylor Wimpey
22	Experian		52	Tesco
23	Fresnillo		53	Unilever
24	Glaxosmithkline		54	Vodafone
25	Glencore		55	Whitbread
26	HSBC			
27	Intercontinental Hotels			
28	Intertek			
29	Johnson Matthey			
30	Kingfisher			

Table A2: Correlations

This table displays correlations between market and tweet features. Market features include daily (log)return, (normalized) trading volume, Parkinson (1980) volatility measure, Bid-ask spreads. Message is the log of number of tweets. Positiveness (Agreement) measures aggregate sentiment (degree of agreement) and is defined in (1) and (2). Bots is the log of number of tweets generated by automatic algorithm. * denotes correlations that are significantly different from 0 at the 1% significance level.

	Return	Log(volume)	Volatility	Bid-ask spreads	Message	Positiveness	Agreement	Bots
Log(volume)	-0.0272*							
High-low range	-0.142*	0.273*						
Bid-ask spreads	0.0153	-0.197*	0.0033					
Message	-0.0061	0.178*	0.0951*	0.004				
Positiveness	0.0189*	-0.110*	-0.0780*	-0.0039	-0.0293*			
Agreement	0.01	-0.162*	-0.0896*	-0.0024	-0.427*	0.603*		
Bots	0.003	0.150*	0.106*	-0.0398*	0.791*	-0.0468*	-0.345*	

Table A3. Differences of high-bot vs. low-bot subsamples

This table reports differences between high versus low bot activities subsamples. The samples covers 55 of FTSE 100 composites from 1st August 2015 to 31st July 2017. High-bot (low-bot) denotes at least (at most) 50% of the total tweets. The last column reports t-statistics from t-tests with the unequal variances assumption. *, **, *** denotes 5%, 1%, 0.1% significance, respectively.

	High-bot	**Low-bot**	**t-stat.**
P/E ratio	23.1077	25.1979	-2.1403**
Obs.	988	23,229	
Earnings per share	76.4773	67.6716	3.5786***
Obs.	994	23,646	
Book value	5.9194	6.8217	-3.8520***
Obs.	994	23,646	
ROE (%)	22.8442	13.6436	12.7685***
Obs.	968	23,224	
Leverage (%)	155.6371	119.8921	6.2090***
Obs.	994	23,646	
Return (%)	-0.14	0.02	-3.3106**
Obs.	994	23,646	
Log(volume)	8.0531	8.3259	-5.6322***
Obs.	967	22,793	
Volatility (%)	1.42	1.71	-9.9680***
Obs.	967	22,793	
Bid-ask spreads (basis points)	5.2594	5.4805	-1.7726
Obs.	967	22,793	
No. of tweets	266.0532	2,100	-25.0616***
Obs.	939	23,646	
Positiveness	0.7415	1.1181	-10.6974***
Obs.	939	23,646	
Agreement	0.4288	0.2769	9.2185***
Obs.	686	23,403	

Table A4: Lagged tweets regressions for returns

This table reports fixed-effects regressions of stock returns by tweets. Dependent variables are (log)return, mean-adjusted return, market-model return. Main independent variables are lagged tweet features. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (tweets generated by automatic algorithm accounts). Bots × Positiveness (Bots × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Return	(2) Return	(3) Mean-adjusted return	(4) Mean-adjusted return	(5) Market-model return	(6) Market-model return
Lagged Positiveness	0.0029	-0.0018	-0.0003	-0.0045	0.0008	-0.0052
	(0.29)	(-0.17)	(-0.03)	(-0.41)	(0.07)	(-0.46)
Lagged Message	0.0387	0.0236	0.0535**	0.0319	0.0103	0.0050
	(1.92)	(1.03)	(2.65)	(1.39)	(0.51)	(0.22)
Lagged Agreement	-0.0010	0.0031	-0.0029	0.0017	-0.0041	-0.0002
	(-0.10)	(0.30)	(-0.29)	(0.16)	(-0.40)	(-0.02)
Lagged FTSE 100 return	0.0105	0.0108	0.0115	0.0117	0.0063	0.0071
	(1.17)	(1.20)	(1.28)	(1.29)	(0.77)	(0.86)
Lagged Bots		0.0085		0.0166		0.0121
		(0.66)		(1.29)		(0.94)
Bots × Positiveness		-0.0126		-0.0112		-0.0194
		(-1.19)		(-1.06)		(-1.79)
Bots × Agreement		0.0058		0.0072		0.0078
		(0.60)		(0.75)		(0.79)
Observations	24,089	23,739	24,089	23,739	24,089	23,739
R^2	0.002	0.002	0.001	0.001	0.001	0.001

Table A5: Lagged tweets regressions for volatility, trading volume, bid-ask spreads

This table reports fixed-effects regression of stock volatility, trading volume, bid-ask spreads by tweets. Dependent variables are volatility measure i.e. Parkinson (1980) intraday high-low range, trading volume, and bid-ask spreads. Main independent variables are lagged tweet features. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (tweets generated by automatic algorithm accounts). Bots × Positiveness (Bots × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Volatility	(2) Volatility	(3) Volume	(4) Volume	(5) Bid-Ask	(6) Bid-Ask
Lagged Positiveness	-0.0332**	-0.0238*	-0.0100***	-0.0077*	-0.0132	-0.0091
	(-3.21)	(-2.16)	(-3.44)	(-2.52)	(-1.67)	(-1.08)
Lagged Message	0.2623***	0.1875***	0.0781***	0.0635***	-0.0278	-0.0564**
	(13.19)	(8.35)	(14.60)	(10.34)	(-1.72)	(-3.04)
Lagged Agreement	-0.0113	-0.0053	-0.0037	-0.0026	-0.0143	-0.0126
	(-1.15)	(-0.54)	(-1.20)	(-0.83)	(-1.51)	(-1.32)
Lagged FTSE 100 return	-0.0471***	-0.0470***	-0.0224***	-0.0215***	0.0226***	0.0222***
	(-5.89)	(-5.85)	(-10.01)	(-9.53)	(3.78)	(3.70)
Lagged Bots		0.0903***		0.0086*		0.0361**
		(7.63)		(2.18)		(2.97)
Bots × Positiveness		0.0206		0.0001		0.0175
		(1.89)		(0.02)		(1.83)
Bots × Agreement		0.0286**		0.0174***		-0.0021
		(3.13)		(5.44)		(-0.21)
Observations	23,214	22,896	23,214	22,896	23,214	22,896
R^2	0.219	0.224	0.915	0.915	0.226	0.226

Table A6: Regressions of returns by bot proportion

This table reports fixed-effects regressions of stock returns by tweets. Dependent variables are (log)return, mean-adjusted return, market-model return. Main independent variables are tweet features. Positiveness and Agreement are defined in equations (1) and (2). Message is the log of number of tweets. Bot presence is the proportion of tweets generated by automatic algorithm accounts. Bot presence × Positiveness (Bot presence × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Return	(2) Return	(3) Mean-adjusted return	(4) Mean-adjusted return	(5) Market-model return	(6) Market-model return
Positiveness	0.0239*	0.0128	0.0207*	0.0011	0.0251*	0.0014
	(2.49)	(0.90)	(2.15)	(0.07)	(2.17)	(0.08)
Message	-0.0324	-0.0277	-0.0174	-0.0096	-0.0308	-0.0244
	(-1.33)	(-1.13)	(-0.71)	(-0.39)	(-1.06)	(-0.84)
Agreement	-0.0088	-0.0089	-0.0107	-0.0082	-0.0102	-0.0034
	(-1.00)	(-0.62)	(-1.21)	(-0.57)	(-0.96)	(-0.19)
FTSE 100 return	0.5495***	0.5493***	0.5491***	0.5489***	0.0333***	0.0331***
	(69.12)	(69.21)	(69.25)	(69.35)	(3.56)	(3.55)
Bot presence		0.0163		0.0256*		0.0206
		(1.54)		(2.43)		(1.61)
Bot presence × Positiveness		-0.0127		-0.0229		-0.0284
		(-0.70)		(-1.27)		(-1.30)
Bot presence × Agreement		-0.0023		-0.0006		0.0051
		(-0.14)		(-0.04)		(0.26)
Observations	24,089	24,089	24,089	24,089	24,089	24,089
R^2	0.304	0.304	0.302	0.303	0.002	0.002

Table A7: Regressions of volatility, trading volume, bid-ask spreads by bot proportion

This table reports fixed-effects regression of stock volatility, trading volume, bid-ask spreads by tweets. Dependent variables are volatility measure i.e. Parkinson (1980) intraday high-low range, trading volume, and bid-ask spreads. Main independent variables are tweet features. Positiveness and Agreement are defined in equations (1) and (2). Message is the log of number of tweets. Bot presence is the proportion of tweets generated by automatic algorithm accounts. Bot presence × Positiveness (Bot presence × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Volatility	(2) Volatility	(3) Volume	(4) Volume	(5) Bid-Ask	(6) Bid-Ask
Positiveness	-0.0387***	-0.0685***	-0.0027	-0.0013	-0.0143	-0.0041
	(-3.37)	(-4.10)	(-0.96)	(-0.23)	(-1.70)	(-0.24)
Message	0.3841***	0.4213***	0.0739***	0.0840***	-0.0162	-0.0090
	(14.50)	(16.21)	(12.32)	(13.77)	(-0.91)	(-0.48)
Agreement	-0.0158	0.0083	-0.0166***	-0.0083	-0.0005	0.0022
	(-1.59)	(0.53)	(-5.38)	(-1.46)	(-0.05)	(0.12)
FTSE 100 return	-0.0871***	-0.0884***	-0.0130***	-0.0133***	0.0090	0.0087
	(-8.72)	(-8.88)	(-6.38)	(-6.57)	(1.82)	(1.76)
Bot presence		0.0822***		0.0163***		0.0098
		(6.79)		(4.39)		(0.85)
Bot presence × Positiveness		-0.0362		0.0012		0.0120
		(-1.60)		(0.19)		(0.66)
Bot presence × Agreement		0.0193		0.0084		0.0026
		(1.04)		(1.44)		(0.15)
Observations	23,242	23,242	23,242	23,242	23,242	23,242
R^2	0.239	0.242	0.915	0.915	0.227	0.227

Table A8: Regressions of returns by lagged bot proportion

This table reports fixed-effects regressions of stock returns by tweets. Dependent variables are (log)return, mean-adjusted return, market-model return. Main independent variables are lagged tweet features. Positiveness and Agreement are defined in equations (1) and (2). Message is the log of number of tweets. Bot presence is the proportion of tweets generated by automatic algorithm accounts. Bot presence × Positiveness (Bot presence × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Return	(2) Return	(3) Mean-adjusted return	(4) Mean-adjusted return	(5) Market-model return	(6) Market-model return
Lagged Positiveness	0.0029	0.0116	-0.0003	-0.0002	0.0008	0.0024
	(0.29)	(0.63)	(-0.03)	(-0.01)	(0.07)	(0.13)
Lagged Message	0.0387	0.0403	0.0535**	0.0583**	0.0103	0.0129
	(1.92)	(1.96)	(2.65)	(2.82)	(0.51)	(0.62)
Lagged Agreement	-0.0010	0.0024	-0.0029	0.0031	-0.0041	0.0038
	(-0.10)	(0.13)	(-0.29)	(0.17)	(-0.40)	(0.20)
Lagged FTSE 100 return	0.0105	0.0105	0.0115	0.0113	0.0063	0.0063
	(1.17)	(1.16)	(1.28)	(1.26)	(0.77)	(0.76)
Lagged Bot presence		-0.0030		0.0064		-0.0008
		(-0.24)		(0.51)		(-0.07)
Lagged Bot presence × Positiveness		0.0100		-0.0004		0.0010
		(0.48)		(-0.02)		(0.05)
Lagged Bot presence × Agreement		0.0048		0.0066		0.0097
		(0.26)		(0.36)		(0.50)
Observations	24,089	24,089	24,089	24,089	24,089	24,089
R^2	0.002	0.002	0.001	0.001	0.001	0.001

Table A9: Regressions of volatility, trading volume, bid-ask spreads by lagged bot proportion

This table reports fixed-effects regression of stock volatility, trading volume, bid-ask spreads by tweets. Dependent variables are volatility measure i.e. Parkinson (1980) intraday high-low range, trading volume, and bid-ask spreads. Main independent variables are lagged tweet features. Positiveness and Agreement are defined in equations (1) and (2). Message is the log of number of tweets. Bot presence is the proportion of tweets generated by automatic algorithm accounts. Bot presence × Positiveness (Bot presence × Agreement) is the interaction between bot activities and Positiveness (Agreement). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Volatility	(2) Volatility	(3) Volume	(4) Volume	(5) Bid-Ask	(6) Bid-Ask
Lagged Positiveness	-0.0332**	-0.1087***	-0.0100***	-0.0116*	-0.0132	-0.0068
	(-3.21)	(-5.53)	(-3.44)	(-1.99)	(-1.67)	(-0.41)
Lagged Message	0.2623***	0.2994***	0.0781***	0.0830***	-0.0278	-0.0164
	(13.19)	(15.25)	(14.60)	(15.18)	(-1.72)	(-0.97)
Lagged Agreement	-0.0113	0.0344	-0.0037	0.0085	-0.0143	-0.0152
	(-1.15)	(1.84)	(-1.20)	(1.44)	(-1.51)	(-0.85)
Lagged FTSE 100 return	-0.0471***	-0.0483***	-0.0224***	-0.0225***	0.0226***	0.0221***
	(-5.89)	(-6.06)	(-10.01)	(-10.07)	(3.78)	(3.70)
Lagged Bot presence		0.0869***		0.0026		0.0244*
		(7.21)		(0.71)		(2.24)
Lagged Bot presence × Positiveness		-0.0922***		-0.0032		0.0084
		(-3.87)		(-0.50)		(0.48)
Lagged Bot presence × Agreement		0.0424*		0.0144*		-0.0038
		(2.28)		(2.45)		(-0.22)
Observations	23,214	23,214	23,214	23,214	23,214	23,214
R^2	0.219	0.223	0.915	0.915	0.226	0.226

Table A10: Intraday regressions based on pre-trading-hour tweets

This table reports fixed-effects regressions of stock returns, high-low range, trading volume by tweets based on 5-minute data from February to July 2017. Dependent variables are (log)return, mean-adjusted return, market-model return, volatility, (normalized) trading volume. Main independent variables are features of tweets collected during 5-minute intervals. We exclude tweets during trading hours 8:00 - 16:30. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (bot-tweets). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Return	(2) Return	(3) Mean-adj. return	(4) Mean-adj. return	(5) Mrk-model return	(6) Mrk-model return	(7) Volatility	(8) Volatility	(9) Volume	(10) Volume
Positiveness	0.0011	-0.0093	-0.0002	-0.0078	-0.0012	-0.0034	0.0088***	0.0116	0.0154*	0.1073***
	(0.31)	(-0.37)	(-0.06)	(-0.31)	(-0.35)	(-0.13)	(5.46)	(1.21)	(2.18)	(3.55)
Message	-0.0012	-0.0011	-0.0010	-0.0007	0.0000	0.0004	0.0389***	0.0319***	0.1056***	0.0969***
	(-0.56)	(-0.42)	(-0.46)	(-0.26)	(0.02)	(0.15)	(45.91)	(28.68)	(29.06)	(17.35)
Agreement	0.0011	0.0170	0.0026	0.0132	0.0048	0.0114	-0.0138***	0.0284	-0.0085	-0.0724
	(0.18)	(0.31)	(0.45)	(0.24)	(0.80)	(0.21)	(-5.05)	(1.12)	(-0.56)	(-0.69)
FTSE 100 return	83.3951***	83.4131***	83.4966***	83.5136***	-23.4718***	-23.4598***	4.7691***	4.6604***	15.3858***	15.1201***
	(32.65)	(32.62)	(32.62)	(32.59)	(-8.51)	(-8.50)	(4.26)	(4.15)	(5.00)	(4.90)
Bots		0.0004		-0.0001		-0.0007		0.0111***		0.0087
		(0.08)		(-0.02)		(-0.13)		(5.32)		(0.98)
Bots × Positiveness		-0.0016		-0.0012		-0.0003		0.0003		0.0145**
		(-0.43)		(-0.31)		(-0.08)		(0.19)		(3.02)
Bots × Agreement		0.0024		0.0016		0.0010		0.0070		-0.0092
		(0.31)		(0.20)		(0.12)		(1.87)		(-0.60)
N	73,117	73,117	73,117	73,117	73,117	73,117	73,117	73,117	73,117	73,117
R^2	0.123	0.123	0.190	0.190	0.085	0.085	0.205	0.207	0.704	0.704

Table A11: Lagged 5-minute regressions based on pre-trading-hour tweets

This table reports fixed-effects regressions of stock returns, high-low range, trading volume by tweets based on 5-minute data from February to July 2017. Dependent variables are (log)return, mean-adjusted return, market-model return, volatility, (normalized) trading volume. Main independent variables are features of tweets collected during the previous 5-minute intervals. We exclude tweets during trading hours 8:00 - 16:30. Positiveness and Agreement are defined in equations (1) and (2). Message (Bots) is the log of number of tweets (bot-tweets). T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

	(1) Return	(2) Return	(3) Mean-adj. return	(4) Mean-adj. return	(5) Mrk-model return	(6) Mrk-model return	(7) Volatility	(8) Volatility	(9) Volume	(10) Volume
Lagged Positiveness	0.0005	-0.0172	-0.0007	-0.0161	-0.0030	-0.0104	0.0057***	-0.0199**	0.0006	-0.0473
	(0.21)	(-1.60)	(-0.28)	(-1.48)	(-1.36)	(-1.02)	(4.54)	(-3.08)	(0.08)	(-1.53)
Lagged Message	0.0033**	0.0027	0.0035**	0.0031	0.0001	0.0002	0.0172***	0.0174***	0.0563***	0.0590***
	(2.73)	(1.55)	(2.89)	(1.75)	(0.07)	(0.10)	(26.50)	(18.69)	(15.34)	(10.62)
Lagged Agreement	0.0030	-0.0399	0.0047	-0.0411	0.0062	-0.0551	-0.0168***	0.0725***	-0.0249	0.1628
	(0.60)	(-1.10)	(0.92)	(-1.12)	(1.27)	(-1.57)	(-6.47)	(3.56)	(-1.64)	(1.45)
Lagged FTSE 100 return	1.4898	1.4710	1.5457	1.5260	-3.0964**	-3.1227**	1.7542**	1.8335**	6.1082*	6.3113*
	(1.32)	(1.31)	(1.37)	(1.35)	(-2.90)	(-2.93)	(2.65)	(2.77)	(2.06)	(2.13)
Lagged Bots		0.0040		0.0036		0.0023		-0.0006		-0.0057
		(1.37)		(1.23)		(0.85)		(-0.34)		(-0.64)
Lagged Bots × Positiveness		-0.0031		-0.0027		-0.0014		-0.0039***		-0.0071
		(-1.82)		(-1.58)		(-0.89)		(-3.84)		(-1.45)
Lagged Bots × Agreement		-0.0061		-0.0065		-0.0089		0.0133***		0.0277
		(-1.15)		(-1.23)		(-1.75)		(4.46)		(1.69)
N	72,919	72,919	72,919	72,919	72,919	72,919	72,919	72,919	72,919	72,919
R^2	0.001	0.001	0.080	0.080	0.076	0.076	0.106	0.106	0.698	0.698

Table A12: Event study – Response of returns

This table reports the decomposition of cumulative (abnormal) returns in response to abnormal increases in different tweets. An abnormal increase in tweets satisfies all the following three conditions: (i) in the top 5% of the empirical distribution of daily changes in each firm; (ii) relative change is larger than 100%; (iii) absolute change is larger than 500 (100 for bot activities). [0], [1], [1,5] report average cumulative changes in percentage points. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

Time windows	All	Positive	Negative	Bots	Bot pos.	Bot neg.	Human	Human pos.	Human neg.
Panel A: Response of returns									
Responses to abnormal increases in tweets									
[0]	-0.148	-0.228	-0.670**	-0.282	-0.121	-0.269	-0.220	-0.360	-0.609*
[1]	0.025	0.143	-0.069	-0.018	0.159	-0.068	0.096	0.218*	0.011
[1,5]	0.479**	0.204	0.663*	0.724**	0.347	0.995**	0.491*	0.634**	0.630
Obs.	751	426	306	361	153	77	459	186	117
Responses to abnormal increases in original tweets									
[0]	-0.042	-0.035	-0.552*	-0.304	0.027	-0.206	0.000	-0.388	-0.561
[1]	0.004	0.036	0.148	-0.023	0.058	-0.079	-0.089	0.064	0.018
[1,5]	0.086	0.218	0.188	0.641**	0.261	0.862*	0.153	0.084	-0.379
Obs.	623	286	187	340	141	51	421	165	85
Responses to abnormal increases in retweets									
[0]	-0.293*	-0.371*	-0.563*	-0.056	0.231	-0.435	-0.248	-0.487**	-0.224
[1]	0.051	0.165	0.076	-0.048	0.171	-0.384	0.132	0.117	0.191
[1,5]	0.640**	0.646**	0.915**	0.421	0.394	0.211	0.735**	0.616**	0.740*
Obs.	518	313	221	179	57	26	372	178	99
Panel B: Response of mean-adjusted returns									
Responses to abnormal increases in tweets									
[0]	-0.134	-0.212	-0.653**	-0.261	-0.125	-0.264	-0.186	-0.359	-0.598*
[1]	0.040	0.161	-0.050	0.005	0.157	-0.059	0.130	0.221*	0.025
[1,5]	0.538***	0.274	0.739**	0.811***	0.337	1.029**	0.630**	0.644**	0.685
Obs.	751	426	306	361	153	77	459	186	117
Responses to abnormal increases in original tweets									
[0]	-0.015	-0.008	-0.529*	-0.288	0.019	-0.194	0.033	-0.381	-0.551
[1]	0.032	0.064	0.174	-0.006	0.053	-0.065	-0.055	0.074	0.030
[1,5]	0.200	0.331	0.284	0.707**	0.237	0.919*	0.291	0.121	-0.341
Obs.	623	286	187	340	141	51	421	165	85
Responses to abnormal increases in retweets									
[0]	-0.277*	-0.359	-0.555*	-0.058	0.203	-0.479	-0.226	-0.487**	-0.227
[1]	0.068	0.179	0.085	-0.048	0.147	-0.426	0.156	0.117	0.190
[1,5]	0.706***	0.700**	0.954**	0.417	0.301	0.059	0.829***	0.615**	0.741*
Obs.	518	313	221	179	57	26	372	178	99

Table A13: Event study – Response of market-model abnormal returns and volatility

This table reports the decomposition of cumulative abnormal returns, volatility changes in response to abnormal increases in different tweets. An abnormal increase in tweets satisfies all the following three conditions: (i) in the top 5% of the empirical distribution of daily changes in each firm; (ii) relative change is larger than 100%; (iii) absolute change is larger than 500 (100 for bot activities). [0], [1], [1,5] report average cumulative changes in percentage points. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

Time windows	All	Positive	Negative	Bots	Bot pos.	Bot neg.	Human	Human pos.	Human neg.
Panel A: Response of market-model abnormal returns									
Responses to abnormal increases in tweets									
[0]	-0.033	-0.073	-0.530*	-0.122	-0.070	-0.317	-0.132	-0.180	-0.457*
[1]	0.008	0.104	-0.061	-0.013	0.155	0.039	0.084	0.180*	0.018
[1,5]	0.320**	0.092	0.311	0.463**	-0.039	0.852**	0.385*	0.300	0.207
Obs.	751	426	306	361	153	77	459	186	117
Responses to abnormal increases in original tweets									
[0]	0.070	0.089	-0.296	-0.131	0.043	-0.297	0.101	-0.166	-0.361
[1]	-0.007	0.049	0.088	-0.043	-0.003	-0.033	-0.065	0.026	0.085
[1,5]	0.074	0.092	0.081	0.395*	0.007	0.552	0.112	-0.209	-0.518
Obs.	623	286	187	340	141	51	421	165	85
Responses to abnormal increases in retweets									
[0]	-0.156	-0.231	-0.438*	0.031	0.294	-0.361	-0.098	-0.307*	-0.151
[1]	0.100	0.158	0.060	-0.101	0.118	-0.136	0.182*	0.101	0.096
[1,5]	0.331*	0.415*	0.453*	0.228	0.140	-0.072	0.447*	0.319	0.317
Obs.	518	313	221	179	57	26	372	178	99
Panel B: Response of volatility									
Responses to abnormal increases in tweets									
[0]	0.616***	0.554***	0.690***	0.436***	0.306*	0.078	0.507***	0.305*	0.392*
[1]	0.273***	0.220***	0.376***	0.201**	0.135	0.099	0.290***	0.090	0.263**
[1,5]	0.524***	0.601***	0.541**	0.612***	0.253	-0.073	0.699***	0.107	0.481
Obs.	718	417	293	352	149	72	448	182	113
Responses to abnormal increases in original tweets									
[0]	0.747***	0.670***	0.702***	0.382**	0.226*	0.116	0.660***	0.429**	0.480*
[1]	0.323***	0.364***	0.412***	0.227***	0.184*	-0.014	0.400***	0.370***	0.380**
[1,5]	0.831***	0.894***	0.910**	0.546**	0.484*	0.125	0.851***	0.600**	0.862*
Obs.	594	281	182	334	137	49	411	164	84
Responses to abnormal increases in retweets									
[0]	0.475***	0.354**	0.443*	0.109	0.010	-0.085	0.371**	0.169	0.040
[1]	0.269***	0.219**	0.337*	-0.032	-0.172*	-0.002	0.256***	0.189*	0.072
[1,5]	0.532**	0.410*	0.527*	-0.149	-0.588*	-0.331	0.523**	0.142	-0.061
Obs.	500	306	212	172	54	23	362	172	95

Table A14: Event study – Response of trading volume and bid-ask spreads

This table reports the decomposition of cumulative changes in (normalized) trading volume, bid-ask spreads in response to abnormal increases in different tweets. An abnormal increase in tweets satisfies all the following three conditions : (i) in the top 5% of the empirical distribution of daily changes in each firm; (ii) relative change is larger than 100%; (iii) absolute change is larger than 500 (100 for bot activities). [0], [1], [1,5] report average cumulative changes in percentage points. *, **, *** denote 5%, 1%, 0.1% significance, respectively.

Time windows	All	Positive	Negative	Bots	Bot pos.	Bot neg.	Human	Human pos.	Human neg.
Panel A: Response of (normalized) trading volume									
Responses to abnormal increases in tweets									
[0]	18.81***	6.05***	15.11***	9.28***	3.77	2.64	14.89***	6.30	8.63
[1]	13.27***	10.13***	16.69***	9.44***	4.73	4.83	13.11***	9.45**	10.85*
[1,5]	31.04***	25.67***	41.05***	22.28**	11.44	5.49	32.80***	16.58*	29.90*
Obs.	718	417	293	352	149	72	448	182	113
Responses to abnormal increases in original tweets									
[0]	22.21***	16.97***	12.91**	8.52**	3.07	0.73	18.38***	13.14**	10.72*
[1]	17.30***	13.66***	18.08***	12.28***	7.57*	-0.19	17.18***	17.93***	20.12***
[1,5]	34.45***	28.44***	33.69**	23.30**	21.89*	-1.30	36.02***	36.88***	27.18*
Obs.	594	281	182	334	137	49	411	164	84
Responses to abnormal increases in retweets									
[0]	14.27***	13.36***	9.10*	4.08	-3.12	-3.02	11.72***	9.33**	2.04
[1]	10.49***	7.75**	13.36***	6.51*	-5.45	0.40	9.30***	8.49**	8.60
[1,5]	34.05***	24.40**	44.41***	8.47	-26.02	-0.36	34.30***	17.79*	16.94
Obs.	500	306	212	172	54	23	362	172	95
Panel B: Response of bid-ask spreads									
Responses to abnormal increases in tweets									
[0]	-10.75***	-8.46***	-9.18***	-10.64***	-5.87	-7.41*	-9.83***	-7.13*	-2.18
[1]	-10.84***	-10.40***	-3.51	-4.16	3.37	-7.09	-7.82***	-3.41	2.28
[1,5]	-45.43***	-37.03***	-32.04***	-29.62***	-34.37***	-12.15	-36.50***	-34.02***	-19.880*
Obs.	718	417	293	352	149	72	448	182	113
Responses to abnormal increases in original tweets									
[0]	-10.08***	-7.74**	-10.53***	-13.98***	-10.35**	-10.43*	-9.28***	-9.79**	-8.02*
[1]	-11.58***	-8.34***	-8.49**	-7.26**	1.49	-6.77	-9.01***	-5.10	-6.66
[1,5]	-41.93***	-24.54***	-23.35**	-28.19***	-33.08***	-19.41	-37.76***	-22.54**	-16.73
Obs.	594	281	182	334	137	49	411	164	84
Responses to abnormal increases in retweets									
[0]	-9.32***	-6.17*	-7.26*	-4.11	-7.03	-4.56	-8.17***	-5.33	-5.33
[1]	-7.00***	-7.88**	-4.68	-4.05	-6.05	0.70	-4.59*	-1.50	-2.60
[1,5]	-43.13***	-41.38***	-32.56***	-28.38***	-34.27**	-9.35	-32.84***	-39.67***	-22.92*
Obs.	500	306	212	172	54	23	362	172	95

Appendix B

This table presents examples of polarity score of Twitter postings calculated by TextBlob.

Text	Sentiment score
FTSE	
Ivans having a terrible time as Glencore earning slump	-1
EasyJet your new seats are sick	-0.43
Why does Sainsburys want to buy Argos	0
CocaCola is the first fortune 500 company to replenish all water used globally	0.125
Beautiful Burberry women's wool cashmere peacoat jacket sz US 8 EU 42 look	1